



Munich Personal RePEc Archive

Commuting and internet traffic congestion

Berliant, Marcus

Washington University in St. Louis

9 March 2017

Online at <https://mpra.ub.uni-muenchen.de/77378/>

MPRA Paper No. 77378, posted 10 Mar 2017 06:02 UTC

Commuting and Internet Traffic Congestion*

Marcus Berliant^{†‡}

March 2017

Abstract

We examine the fine microstructure of commuting in a game-theoretic setting with a continuum of commuters. Commuters' home and work locations can be heterogeneous. A commuter transport network is exogenous. Traffic speed is determined by link capacity and by local congestion at a time and place along a link, where local congestion at a time and place is endogenous. The model can be reinterpreted to apply to congestion on the internet. We find sufficient conditions for existence of equilibrium, that multiple equilibria are ubiquitous, and that the welfare properties of morning and evening commute equilibria differ.

JEL numbers: L86, R41 Keywords: Commuting; Congestion externality; Efficient Nash equilibrium

*The author is grateful to David Boyce, whose address at the Regional Science Association International North American Meetings in Savannah incited the author (who was also his discussant) to write this paper. He is grateful to the Helen R. Whiteley Center for its kind support and hospitality. Alex Anas, Richard Arnott, Gilles Duranton, Jan Eeckhout, Ignacio Esponda, Mogens Fosgerau, Jonathan Hall, Eren Inci, Kamhon Kan, T. John Kim, Lewis Kornhauser, Bill Neilson, Kirill Pogorelskiy, Tony Smith, Matt Turner, Erik Verhoef and Ping Wang contributed interesting comments. The author is grateful to seminar audiences at the Public Economic Theory meetings in Galway, the Summer Meetings of the Econometric Society in Tokyo, the Kuhmo-Nectar Conference in Stockholm, the Regional Science Association International meetings in Miami, the Institute of Economic Research at Kyoto University, the University of Tokyo, Academia Sinica, Washington University in St. Louis, and the Seventh Kansas Workshop on Economic Theory for comments. The author remains responsible for any opinions or errors herein.

[†]Department of Economics, Washington University, Campus Box 1208, 1 Brookings Drive, St. Louis, MO 63130-4899 USA. Phone: (314) 935-8486, Fax: (314) 935-4156, e-mail: berliant@wustl.edu

[‡]Institute of Economic Research, Kyoto University and California Institute of Technology.

1 Introduction

1.1 Motivation

Commuting is a ubiquitous feature of the urban economy. Although the classic literature has answered some basic questions in the field, such as whether equilibrium commuting patterns are generally efficient, surprisingly some very important questions remain open. If cars can catch up with each other, what can we say about endogenous equilibrium congestion? Do models without an explicit continuous time clock give us an accurate picture of traffic, in the sense that they can approximate behavior in a truly dynamic model? In contrast with most of the literature, our model says that multiple equilibria are to be expected. Can traffic be improved simply by equilibrium selection rather than congestion pricing? Are there distinct welfare differences between Nash equilibria for commuting from home to work in contrast with commuting from work to home on the same network? The last question is perhaps the most important.¹ Our fine microstructure allows us to examine these differences.

There is an important application of our model to traffic and congestion on the internet. Instead of cars, packets of information move over the network, each with a given origin and destination. Both positive and normative questions concerning route choice and departure time can be addressed with our model.² Interestingly, both the car and internet congestion literatures began with discrete models (at different times), and eventually moved to continuous flow models for tractability reasons.

The economic models employed in the commuting literature are often very special and unrealistic; a literature review will be provided in the next subsection. One class of models features identical commuters, a very simple network structure (for example a home, a workplace and one link between them), and an exogenous bottleneck that results in queuing of traffic. It is not known to what degree the results derived in the literature rely on these or other strong simplifying assumptions that generally provide a reduced form viewpoint. In contrast, we study a new class of more natural models that allows arbitrary heterogeneity in both commuters and network structure (for example allowing cross-commuting), where congestion is endogenous and traffic slows in response

¹For those who wish to skip ahead to this item, please see Example 4 and Theorem 3.

²In general, one user will send out many packets. However, if these represent a negligible proportion of the total number of packets, coordination of the strategy choices for these packets is the same as no coordination for our purposes.

to congestion relative to road capacity. In the last subsection of the introduction, we will provide simple examples that display the contrast between the existing literature and our class of models.

There are important differences in implications between our framework and the existing literature, mainly due to the detailing of fine microstructure in our work. The reduced forms, such as an exogenous congestion function, used elsewhere are generally not supported by this microstructure, leading to different results. Our model employs a microfounded, evolving congestion concept that is suggested by the transportation engineering literature. Thus, the conditions sufficient for existence of equilibrium are markedly different. As we shall see in the examples, it is quite natural to have multiple equilibria in our framework, whereas the goal of the existing literature is often to prove that equilibrium is unique. Finally, as we shall illustrate, equilibria in our framework are qualitatively very different from those derived in the rest of the literature, mainly due to the fine microstructure.

1.2 Five Related Literatures

Before proceeding to our examples and analysis, we discuss the basic literature on congestion. We divide this literature into 5 components: the transportation economics literature, the game-theoretic literature on congestion externalities, the transportation engineering literature, the mathematics of conservation laws, and the electrical engineering literature on internet congestion.³ We discuss these in turn. Our work is at the junction of all of these literatures. In contrast with our work, the first two literatures tend not to study dynamic micro behavior along roads. The second two literatures take individual behavior as fixed, so the models are mechanical. The last literature tends not to examine Nash equilibrium, but rather other positive or normative ideas.

The older literature on transportation economics deals with models with no time clock or with just one route or bottleneck where traffic queues. Beckmann et al. (1956) provide a model of rush hour where traffic flows are constant. They analyze optimum and equilibrium in a stylized model with no explicit time clock, but with a representative commuter. Vickrey (1963, 1969) provided the classical analysis of congestion externalities, pricing, and infrastructure investment. Arnott et al. (1993) examine primarily welfare under various pricing schemes when there is only one route or bottleneck, but allow elastic

³These literatures tend not to cite each other, rendering literature reviews labor-intensive and occasionally puzzling, due to terminological differences.

trip demand and use continuous time. Traffic does not slow down due to congestion, but rather queues at a bottleneck with limited capacity. In their conclusions (p. 177), they note: “In the context of rush hour traffic congestion, for example, models should be developed which derive hypercongestion (traffic-jam situations) from driving behavior, solve for equilibrium on a congested network, and account for heterogeneity among users...” This is what we attempt.

The contemporary literature on transportation economics uses the terminology “dynamic traffic assignment problem” for the kind of model we shall construct. Merchant and Nemhauser (1978) initiate the modern literature by proposing a discrete time model with a single destination node where events in a link of the transport network at a given time, namely the number of cars entering the link, the cost of traversing the link as a function of traffic, and the number of cars exiting a link as a function of traffic, are all exogenous black boxes. They provide an example and examine algorithms for finding a social optimum. Ross and Yinger (2000) embed a model of point congestion similar to ours in a classic urban monocentric city model with both land consumption and a symmetric radial road network. This is similar to a simple network with only one commuting corridor. Traffic flow is continuous but not necessarily smooth. They show that the only equilibrium in a general urban equilibrium version of a commuting model with continuous departure times and flow congestion but no bottlenecks is an unreasonable one with a never ending rush hour. As we shall explain below, by allowing a large but finite number of departure times and randomizing departures over small intervals between these discrete departure times, with some effort we can overcome these difficulties. In our context, traffic flow might not be continuous. Konishi (2004) considers existence, uniqueness and efficiency of Nash equilibrium primarily in a static model but also in a discrete time dynamic model with a simple network, employing Schmeidler’s (1973) theorem⁴ as we do. He uses bottlenecks whereas we use speed reductions resulting from congestion. Konishi’s work is quite complementary to ours, as we are not concerned with the issues he addresses, namely existence of equilibrium in static models with a finite number of commuters, conditions sufficient for uniqueness of equilibrium in static models with a continuum of commuters, and existence and uniqueness of equilibrium in dynamic models of simple networks with exogenous bottlenecks.

⁴To apply Schmeidler’s work to obtain Nash equilibrium in pure strategies, it is important that the set of pure strategies be finite. In our model, the interpretation is that the set of departure time strategies is finite.

An independent, modern literature in transportation economics examines necessary conditions at a Nash equilibrium for the dynamic traffic assignment problem. Heydecker and Addison (2005) consider what happens along a link as a black box, and derive such a condition. Of course, if such a black box is made more specific, the necessary condition can be refined. Zhang and Zhang (2010) use a bottleneck model and obtain a more specialized condition.

In their survey, de Palma and Fosgerau (2011, p. 208) conclude: “The extension of the dynamic model to large networks remains a difficult problem. So far, existence and uniqueness of equilibrium have not been established (in spite of many attempts).”

The game-theoretic literature on externalities, for example Sandholm (2001), has the potential to be useful in our context. However, the strong symmetry assumptions used, that yield strong and interesting conclusions, exclude almost all of the games of interest to us. For example, they exclude the simple special case of our model where there are two nodes called home and work with one link between them, but two departure times. Hofbauer and Sandholm (2007) study congestion games with a continuum of players, but their assumptions on congestion rule out the type of dynamic micro-interaction along a link that is the focus of our work. Sandholm (2007) considers an evolutionary approach to setting optimal tolls in the case where there is a *finite* number of *identical* commuters (so they have the same home and work locations) modified by an idiosyncratic preference component, without the symmetry assumption but with further structure on the evolutionary process.⁵ Hu (2010) explores Nash equilibrium with continuous departures for a single commuting corridor for one morning rush hour. It is shown that with a specific dynamic for equilibrium selection, the equilibrium exists and is unique. As we shall illustrate in the last subsection of the introduction, multiple equilibria are quite natural in models of commuting.

Naturally, the transportation engineering literature is concerned more with practical traffic issues than with the questions we pose; see, for example, Daganzo (2008). Typically this literature takes the behavior of individuals, namely their choice of routes and departure times, as exogenous. Thus, Nash equilibrium is not studied.⁶ For example, Zhu and Marcotte (2000) use pre-

⁵It is also interesting to inquire how tolls would be implemented in practice in these models, since in theory the toll is based on the overall strategy chosen, namely the route and/or departure time. Would toll booths along the route be able to implement this?

⁶For example, the first appearance of a utility function in Daganzo (2008) is at the bottom of p. 315. The body of the book ends at the top of p. 319.

determined (exogenous) departure times. The closest relative to our model in this literature is the cell transmission version of the Lighthill-Whitman-Richards (LWR) model; see Daganzo (2008) section 4.4.6. There are some important differences. First, the LWR model takes departures as exogenous and possibly smooth, whereas we do not. Second, like most models of traffic, the LWR model employs queues or bottlenecks when there is congestion. In contrast, we assume that traffic slows as a function of traffic density. These two important differences express themselves as differences in the equilibrium behavior of the models.⁷ More recent examples include Han et al (2013)⁸ and Han et al (2015).

Turning next to the mathematics literature, the topics we consider here typically fall into two literatures. In essence, our mathematical problem on one link boils down to a conservation law coupled with a discontinuous differential equation. Even with just one link between an origin and destination with *exogenous* departure times and homogeneous commuters, existence and uniqueness of the resulting traffic pattern is a difficult question that requires interesting assumptions and techniques to resolve. A major issue is the existence and uniqueness of behavior of the system when the initial conditions can be discontinuous. This is important to us, as we don't want to place restrictions on the joint behavior of individuals when we eventually consider Nash equilibrium. The mathematics were introduced in Bressan (2000, chapter 6) and Garavello and Piccoli (2006); that work is based on Bressan (1988) and Bressan and Shen (1998).⁹ The key paper for our purposes is the seminal work of Strub and Bayen (2006), who remark in their conclusions (p. 564), "However,

⁷In other parts of the transportation engineering literature, existence and uniqueness of Nash equilibrium is studied in the context of a bottleneck model, using an S-shaped wish curve (defining ideal bottleneck exit times). In these models, it is unclear what happens if an atom of commuters arrives at the bottleneck at the same time, or if the fragile condition of an S-shaped wish curve is violated - the complement appears to be open and dense in the set of wish curves.

⁸In addition to queues, this work also features a highly non-standard notion of Nash equilibrium.

⁹Although the motivation for Bressan (2000) is the simple traffic problem with one home location, one work location, and one link, the mathematical problem solved in this book is *different* from the economic problem that motivates it. This will cause us some headaches. In particular, the initial condition used in the book is the traffic at various locations along the link at time 0, trivially 0 in our model. Traffic is not allowed to enter the link after time 0. We are much more interested in boundary conditions that, for an arbitrary time, give the traffic entering a link at *location* 0. Nevertheless, the mathematics introduced in this book is very useful.

results are still lacking in order to generalize our approach to a real highway network. For such a network, PDEs are coupled through boundary conditions, which makes the problem harder to pose.”¹⁰ Once we have introduced notation and concepts, we shall remark further on both related literature in mathematics and alternative approaches to solving the induced mathematical problem.¹¹ An important contribution of Strub and Bayen (2006) is actually the *definition* of a solution to the mathematical problem of determining flow in the one link system with *exogenous* departures, since there were issues of either existence or uniqueness with many of the previous attempts. The technical difficulties in the literature are partly the result of working with functions of bounded variation with a two dimensional domain: time and distance. The (discontinuous) conservation law tells us that cars are not lost over a link, with initial condition zero cars on the link and boundary conditions corresponding to departure of cars. The conservation law is coupled with a (discontinuous) differential equation that gives progress of a car over the link. An important mathematical problem is relating properties of functions on two dimensions that are of bounded variation to their variation on each dimension separately as well as on a cone. As shown by Bressan (1988) but not discussed in more recent work in this literature, locally bounded variation on this cone is sufficient to solve the discontinuous ordinary differential equation, associated with a conservation law, for progress on a link.

In the end, we are able to embed the more elementary framework of Strub and Bayen (2006) in a model with an arbitrary transport network, heterogeneous commuters and endogenous choice of departure times and routes, examining Nash equilibrium as well as Pareto optimum. Unfortunately, we cannot apply their results directly, but must open up the details of their clever proof.

The final literature related to our work is the literature on internet congestion. Although we interpret our model as traffic on roads for consistency of exposition, it applies as well to packets on links in the internet. A fine survey

¹⁰There are many challenges that we must address to extend their results from one link to many. For example, it is difficult to prove that the link exit density has the same properties as the link entry density, that is used as the entry density for another link. A secondary challenge is that boundary conditions are formulated in terms of density (cars per mile) when they should be formulated in terms of volume (cars per hour). Although we take the proper approach for boundary conditions using volume, the technicalities can be simplified some if we were to use density.

¹¹We note in frustration that much of the literature cited here is motivated by mathematics rather than economics. Beyond Strub and Bayen (2006), there is no result we can apply directly even to the case of two nodes and one link.

of this literature can be found in Jacobsson (2008). Due to the complexity of the discrete model, a continuous model was developed by Kelly et al (1998), forming a foundation for our work. Much of the literature has a focus on exogenous departures and routes, not Nash equilibrium. Other parts of the literature, such as Kelly et al (1998), focus on steady states of the dynamic model with congestion pricing, or what we call a static model with congestion pricing. There is likely an unexplored relationship with potential games, as represented for example in Sandholm (2007).

1.3 Preview

In summary, the main difference between our work and most of the literature is that we use the fine microstructure from transportation engineering and the mathematics of conservation laws to address more macro economic questions. We do not use exogenous departure and route choices, nor do we employ bottlenecks or queues. Instead we allow endogenous choice of departure times and routes, but require that traffic slow down as a function of endogenous congestion on an arbitrary transportation network. To our knowledge, this represents a new class of models of commuting that has fewer black boxes (such as delay functions in the standard literature) and, more importantly, different properties compared with others.

Although the notation used to describe the models formally is burdensome, we will give examples and intuition for the results in addition to the technicalities. We formulate both a static model, where time plays no role, and a dynamic model, where it does play a role. We assume that commuters have an inelastic demand for one trip per day to work. Future work should extend this to elastic demand.

Our results and the outline of the balance of the paper are as follows. Although classical results concerning Nash equilibrium and Pareto optimum are replicated in our context, we highlight novelties. In the next subsection of the introduction, we detail and preview our results with minimal notation by using the simplest example, a network with two nodes and one link where all commuters live at one node and commute to their jobs at the other. In Section 2, we give our notation and specify the general static (timeless) and dynamic models. At this point, we prove classical results in our context, but also find assumptions sufficient for existence of a unique flow of traffic across links over time given a set of initial conditions (corresponding to a fixed strategy profile) in the dynamic model. Moving on to Nash equilibrium, we find conditions

sufficient to prove it exists, and show that it is generally not unique. Section 3 gives our applications. First, we show that the static model cannot be viewed as a reduced form of the dynamic model, where time is explicit. Then we study the welfare properties of Nash equilibrium in the context of a tree network in the dynamic model. Nash equilibrium of the morning commute will generally be inefficient, whereas there exists a Nash equilibrium of the evening commute that is efficient. Finally, Section 4 gives our conclusions. All proofs are contained in an Appendix.

1.4 Example

1.4.1 Our Basic Model

We begin with a simple example to illustrate how the model works and the intuition behind our results. Consider commuters uniformly distributed on the interval $[0, 1]$ with nodes 1 and 2. Each commuter commutes from node 1 to node 2 each day. For simplicity, we only consider the morning rush hour at this juncture. Denote the capacity of the link by $x \in \mathbb{R}_+$. Suppose that the time it takes to travel the link at the speed limit is $t(1, 2) = 1$. In the static model, the travel time is given by 1 if the average number of travellers does not exceed capacity x of the road, and by $\frac{1}{x}$ otherwise. This means that if road link capacity is exceeded, then traffic slows down in proportion to the ratio of excess commuters to capacity, $\max(1, \frac{1}{x})$. For example, if $x = 1/2$, then the travel time for a commuter on the link is 2. There really are no choices here for the commuters or a social planner optimizing efficiency, since the route is fixed and the model is static; there are no departure times to be chosen.

Now consider a dynamic version of the model. Route choice is still fixed, but departure (and consequent arrival) times are a choice variable of the commuters. We model departure times in \mathbb{R}_+ , and we call the required arrival time at the destination node 2 (say 9 AM) $\tau^A \in \mathbb{R}_+$. There is no penalty for arriving at work early, but the penalty for arriving at work late is ∞ . This is mainly for illustration. We shall consider more general penalties for both early and late arrival in the remaining sections. They add some complications.

Again, in this simple model there is no route choice. But there is a choice of departure time. First, we illustrate how, for any choice of departure times by all commuters, the travel time to the destination node 2 can be computed. It is assumed that the latter is minimized by each individual commuter at a Nash equilibrium (given the choices of others), and the social planner maximizes

a utilitarian welfare function that is minus the integral of commuting times subject to the arrival constraint.

The speed of a particular cohort of commuters who depart at the same time is computed as follows. Begin with the local density of commuters on the road at a particular place on the route and at a particular time. This local density at a given place and time is computed as the limit of neighborhoods on the road of total (measure of) commuters in the neighborhood divided by the one dimensional size of that neighborhood. The limit is taken as the length of the neighborhood goes to zero. The result will be the density of commuters (with respect to distance) at that place and time. Then, as in the static model, traffic slows down in proportion to the ratio of excess commuter density to capacity.

An example will help illustrate. Again consider the commuters uniformly distributed on $[0, 1]$. Suppose that all the commuters at 0 depart at time 0, all the commuters at 1 depart at time 1, and so forth. Set the arrival time $\tau^A = 2$. We compute traffic speeds (in this case, the arrival time constraint will not bind). With these departure times, when road capacity is high so that $x \geq 1$, then capacity does not bind. The unit interval of commuters moves from origin to destination at full speed and perfect synchrony, and the local density of traffic is always 1 except for commuters with labels 0 and 1. The density around them is $\frac{1}{2}$ since there is nobody on one side of them (for example the commuters with label 0 have nobody in front of them). But this does not alter their speed, since they are already at the speed limit. In theory, at least, commuters can catch up with those ahead of them (if the ones ahead are travelling slower) and slow themselves down.

What if $x < 1$? We consider two simple patterns. First, suppose that commuters depart exactly as in the preceding paragraph. Set the arrival time $\tau^A = \frac{1}{x} + 1$. Traffic slows down by a factor of $\frac{1}{x}$ relative to the no congestion case; thus, traffic speed for the commuters is uniform at x . It takes $\frac{1}{x}$ time to traverse the link, so the last commuters (labelled 1) reach the destination at $\frac{1}{x} + 1$. The local density of commuters is 1 during the commute. Call this the *congested commuting pattern*.

Now consider the same general departure pattern as in the preceding paragraph, but with commuters labelled 0 beginning travel at time 0, whereas commuters labelled 1 begin their trip at time $\frac{1}{x}$. So the density of commuters departing at any time is x . Set the arrival time $\tau^A = \frac{1}{x} + 1$. Since local density is the same as capacity, all commuters travel at the speed limit. Thus, travel

time for all commuters is 1. Call this the *uncongested commuting pattern*.

These two simple commuting patterns, or strategy profiles, serve to illustrate the computation of local density and speed. Of course, the local density and speed calculations can be much more complicated in, for example, more intricate commuting networks or for more intricate departure patterns. The simple patterns also serve to illustrate the important role played by arrival time. It is rather evident that for the fixed arrival time as specified at $\tau^A = \frac{1}{x} + 1$, these strategy profiles are Nash equilibria. Notice that all commuters reach work by the arrival time τ^A for either pattern, but travel time is longer for the congested commuting pattern. Thus, welfare can differ across dynamic commuting patterns even for this simple example. It is evident that the uncongested commuting pattern Pareto dominates the congested commuting pattern.

1.4.2 The Classical Model with Queues

A crucial comparison is between our model, with endogenous congestion and speed, and the classical models of the literature that use queues. We argue that the equilibrium (or even disequilibrium) behavior of our model is different and much more realistic, illustrated as follows.

First, consider the model detailed previously. For the purpose of comparison, modify the simple example that we have used by setting uncongested travel time $t(1, 2) = \frac{1}{2}$, arrival time $\tau^A = 2$, and capacity $x = \frac{1}{2}$. In this case, the uncongested commuting pattern has departure times uniformly distributed over $[0, 2]$ with density $\frac{1}{2}$. The time cost is $\frac{1}{2}$ for each commuter, but the commuter departing at time 2 arrives late, so this pattern will not be an equilibrium pattern. For the congested commuting pattern, departure times are uniformly distributed over $[0, 1]$ with density 1, so it takes each commuter time 1 to traverse the link. Thus, the last commuter arrives at time 2, and the congested commuting pattern remains an equilibrium in the modified example.

We turn next to a model with queuing. There are many variations on the bottleneck model, particularly in continuous time. For example, Arnott et al. (1993) assume that it takes no time to get from home to a bottleneck, and that after exiting the bottleneck, the commuter immediately arrives at work. The variation we use is closer to our model, and is due to Zhang and Zhang (2010). A link consists of two parts, a main body first and then a queue at the end. The main body has infinite capacity so traffic flows at the speed limit independent of any congestion. The queue or bottleneck at the end of the

main body operates with limited capacity, using a first-in-first-out principle. For our particular example, it takes time $\frac{1}{2}$ for any commuter to traverse the main body (independent of congestion), and the queue allows density $\frac{1}{2}$ to exit the queue at any given time.

What does Nash equilibrium with a queue look like for this example? Everyone leaves as soon as possible (at time 0), arrives at the bottleneck at time $\frac{1}{2}$, and the last commuter leaves the bottleneck at time $\frac{5}{2}$, late for work. Here, we assume that if everyone arrives at the bottleneck at the same time, the order in which they proceed is random. This equilibrium does not resemble at all the one obtained using our model of endogenous congestion.

1.4.3 Comparison of the Static and Dynamic Models

Consider next the comparison of the static with the dynamic model. We return to the simple, basic example used at the beginning of this subsection where uncongested travel time is $t(1, 2) = 1$. The first pattern, the congested commuting pattern, we study for the case $x < 1$ seems to be the analog of the static case, since traffic speed is constricted. But the second, uncongested pattern does not seem to have an analog. Thus, the static and dynamic models have different Nash equilibrium predictions. Moreover, if the dynamic analog of the static equilibrium is the congested commuting pattern, it is Pareto dominated by another pattern present in the dynamic model but disallowed by the static model.

In fact, we can say more. For example, even in the case where the equilibria of the static and dynamic models appear to be the same, if we average congestion for the dynamic model over time and distances on the link, many times and distances have zero commuters and zero congestion. For instance, this happens at distances along the link in our example that the first commuters have not yet reached. So aggregating the equilibrium of the dynamic model this way will not generate the static model equilibrium, since the flows in the dynamic model will appear diluted.

An alternative criterion for comparing the static and dynamic model equilibria is to ask that there be at least one time and a distance along each link such that the equilibrium flows of the models are the same. An example relevant to this idea is to use three identical links in series, so that at any given time in equilibrium commuters in the dynamic model are on at most two of the links, unless the links are operating below capacity. Then there is no time at which the flows on all three links are non-zero simultaneously. In section 3.1

below, we describe how to extend this example so that there is *no* equilibrium of the dynamic model even remotely resembling the equilibrium of the static model.

With the model specified as we have outlined, generally a Nash equilibrium in pure strategies or an optimum might not exist. So in what follows, for the dynamic model, we must simplify the problem. This is accomplished by using a fixed, finite set of possible departure times that divide equally the time scale in the model. When commuters choose a departure time, they are distributed uniformly over the interval with midpoint their chosen departure time, and length equal to the distance between allowable departure times. With this structure, a Nash equilibrium in pure strategies and an optimum exist. Moreover, for our example, the congested and uncongested commuting patterns we have specified are Nash equilibria of the model, and the uncongested commuting pattern is Pareto optimal.

What follows below just makes the ideas behind our simple examples formal and general, for instance allowing an arbitrary commuting network where commuters have various different origins and destinations.

2 The Commuting Model

Readers who wish to understand the content of the work through examples only can focus on Examples 1-4 below and then skip to section 4.

2.1 The Static Model: Equilibrium and Optimum

Here we lay out the details of a game with an atomless measure space (continuum) of players; a finite set of nodes at which the players live, or to which they commute, or through which they commute; and a finite set of transport links between the nodes with exogenous capacity.

To begin, the measure space of commuters is given by (C, \mathcal{C}, μ) where C is the set of commuters, \mathcal{C} is a σ -algebra on C , and μ is a positive, non-atomic measure.¹² We assume that singletons of the form $\{c\}$ for $c \in C$ are in \mathcal{C} ; that for all $c \in C$, $\mu(\{c\}) = 0$; and $0 < \mu(C) < \infty$.

The origins and destinations in the commuting network are given by a finite set of *nodes*, denoted by $m, n = 1, 2, \dots, N$. Let $\mathcal{N} = \{1, 2, \dots, N\}$. The

¹²Skorokhod's theorem implies that we could without loss of generality restrict attention to the unit interval with Lebesgue measure.

commuting network itself is given by a finite set of links between nodes. The capacity of any direct link (with no intermediate nodes) between nodes m and n is given by $x_{mn} \in [0, \infty]$, whereas $x_{nn} = \infty$. If a direct link between nodes m and n does not exist, then $x_{mn} = 0$.

What remains is to specify the strategies and payoffs of the commuters. In the static game, there is no choice of time of departure or arrival. There is only route choice. We assume that each commuter has a fixed origin node and a fixed destination node, with inelastic demand for exactly one trip between the origin and destination. Thus, there is an exogenous, measurable *origin map* $O : C \rightarrow \mathcal{N}$ and an exogenous, measurable *destination map* $D : C \rightarrow \mathcal{N}$. Notice that there can be heterogeneity among commuters in origins and destinations. This will create heterogeneity in the reduced form utility functions of the commuters.

Let π_k be the map that projects a vector onto its coordinate k . A *route*, denoted by r , is a vector of length no less than 2 but no more than N ,¹³ Next we define the set of all routes:

$$\begin{aligned}\mathfrak{R}^l &= \{r \in \mathcal{N}^l \mid \text{for } i = 1, 2, \dots, l-1, x_{\pi_i(r)\pi_{i+1}(r)} > 0\} \\ \mathfrak{R} &\equiv \bigcup_{l=2}^{\infty} \mathfrak{R}^l\end{aligned}$$

To avoid trivial situations, we assume that if there is a positive measure of commuters with a particular origin and destination, that there is some route between the nodes. A *commuting length map* is a measurable map $l : C \rightarrow \{2, 3, \dots\}$. A *commuting route structure* is a pair (l, R) where l is a commuting length map and R is a measurable map $R : C \rightarrow \mathcal{N}$ such that for $i = 1, 2, \dots, l-1$, $x_{\pi_i(R(c))\pi_{i+1}(R(c))} > 0$, and almost surely for $c \in C$, $\pi_1(R(c)) = O(c)$ and $\pi_{l(c)}(R(c)) = D(c)$.

Given a commuting route structure (l, R) , its *flow* $f \in \mathbb{R}_+^{N^2}$ is given by $f(m, n) = \mu(\{c \in C \mid \exists k \in \{1, 2, \dots, l(c) - 1\} \text{ with } \pi_k(R(c)) = m \text{ and } \pi_{k+1}(R(c)) = n\})$ for $m, n = 1, 2, \dots, N$. We assume that the length of the link between nodes m and n is $\lambda(m, n) \geq 0$ for $m, n = 1, 2, \dots, N$. However, if the link is congested, then the travel time increases. For our examples, it increases in proportion to the excess of commuters above capacity, $\frac{f(m, n)}{x(m, n)} \cdot \lambda(m, n)$.¹⁴ For instance, if the number of commuters is twice the capacity of a link, then the

¹³We can generalize this to an arbitrary but finite upper bound on the length of a route at the cost of more complicated notation.

¹⁴There is an issue of normalization here, namely whether f is divided by λ or not. In essence, it depends on whether a link that is twice as long is half as congested for the same

travel time is doubled. We ask that the reader bear this special case in mind, since we use it in all of our examples to give concrete intuition.

More generally, we can allow traffic to slow down according to any well-behaved function of the number of commuters at a distance on a link and link capacity. Therefore, we specify the function $v : \mathbb{R}_+ \times \mathbb{R}_+ \rightarrow \mathbb{R}_{++}$ where $v(f, x)$ is the speed of traffic with density f on a link with capacity x . We assume that for fixed x , v is continuous and non-increasing in f . For our examples, $v(f, x) \equiv \min\left\{\frac{x}{f}, 1\right\}$.

The *time cost of a commuting structure* (l, R) for commuter c is

$$\theta(l, R, c) = \sum_{n=1}^{l(c)-1} \frac{\lambda(\pi_n(R(c)), \pi_{n+1}(R(c)))}{v(f(\pi_n(R(c)), \pi_{n+1}(R(c))), x(\pi_n(R(c)), \pi_{n+1}(R(c))))} \quad (1)$$

Thus, $-\theta$ is the objective or payoff function for each commuter. The *utilitarian welfare function for the static model* is

$$U(l, R) = - \int_C \theta(l, R, c) d\mu(c)$$

A *Nash equilibrium of the static model* is a commuting structure (l, R) such that almost surely for $c \in C$, there is no route r of length v for commuter c such that

$$\theta(l, R, c) > \sum_{n=1}^{v-1} \frac{\lambda(\pi_n(R(c)), \pi_{n+1}(R(c)))}{v(f(\pi_n(r), \pi_{n+1}(r)), x(\pi_n(r), \pi_{n+1}(r))))}$$

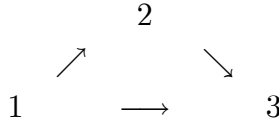
Existence of Nash equilibrium in pure strategies can be proved by applying Schmeidler (1973, Theorems 1 and 2). Rosenthal (1973) proves that a Nash equilibrium in pure strategies exists even when there is a finite number of commuters. Sandholm (2001) shows that equilibrium exists and is unique under additional conditions, primarily that speed is strictly decreasing in link usage f .

Next we prove (informally) that an optimum exists. The problem can easily be reduced to optimization of the utilitarian welfare function over a

number of commuters on the link. This depends on the interpretation of the static model, whether congestion is viewed as a pulse of commuters or whether they are uniformly spread out over the link. In this paper, we take the view that in the model without time, twice as many commuters on a link results in twice the congestion, no matter the length of the link. However, if one takes the view that length of the link matters, the result is simply division of our f by λ , and this makes no essential difference in the the results we obtain. As we show in section 3, interpretation of the static model is difficult.

compact set as follows. Notice first that there is a finite number of types of commuters, defined by their origin-destination pairs. Instead of using route choice for each commuter, employ as control variables the measure of each type following each route. Thus, the social planner controls a finite number of variables in a compact set using a continuous objective, so a maximum is attained.

Example 1: We note that due to the congestion externality, the Nash equilibria are unlikely to be Pareto (or utilitarian) optimal. To see this informally, consider an example with 3 nodes. All commuters travel between nodes 1 and 3. There is a direct route, and an alternate route that runs via node 2. The alternate route takes longer than the direct route for each fixed number of commuters below capacity because it requires a longer distance of travel. For example, each road has capacity 1 and takes 1 unit of time to cross, so the longer route uses 2 units of time when running below capacity, whereas the shorter route takes 1 unit of time when running below capacity.



Suppose that there is measure $\frac{5}{2}$ of commuters. A Nash equilibrium of this model has the direct route running above capacity, with measure 2 commuters using it for a total travel time of 2, and the indirect route running below capacity (.5 measure, with a total travel time of 2) such that the travel time to work for each commuter is the same. To create a Pareto improvement over the Nash equilibrium, simply move some commuters (say measure .5) from the direct to the indirect route. The travel time on the indirect route (namely 2) is the same as at the Nash equilibrium, even for the commuters switched to that route, whereas the travel time for those on the direct route decreases (to 1.5).

2.2 The Dynamic Model: Equilibrium and Optimum

The basics of the dynamic model are the same as those for the static model. To differentiate the notation, we will add “dynamic” to the names and add time τ as an argument of functions. In the dynamic model, each commuter chooses both a departure time (from their origin node) and a route. Routes were discussed in the previous subsection. We allow a commuter to depart at

any time $\tau^d \in [0, T]$. As we shall see shortly, it is important that this set be bounded.

A *dynamic commuting route structure* is a triple (τ^d, l, R) where $\tau^d : C \rightarrow [0, T]$ is a measurable function giving departure times for all commuters, l is a commuting length map and R is a measurable map $R : C \rightarrow \mathcal{N}$ such that almost surely for $c \in C$, $\pi_1(R(c)) = O(c)$ and $\pi_{l(c)}(R(c)) = D(c)$.

At this juncture, there is an issue concerning the detail in which we model congestion on each link in the dynamic model. It varies in the literature we have cited. The simplest way to model this is to look only at average congestion on a link. More complicated is to assume that as traffic ebbs and flows, the congestion at the end of the link determines traffic speed on the entire link. The most detailed model allows cars to catch up with each other over the course of a link. We use this most detailed model, but assume that link capacity is constant across the link. This is without loss of generality, provided that capacity changes only a finite number of times on a link. In that case, we just add more nodes and links with different capacities in series.

We shall define commuter progress from origin to destination through a differential equation in distance. But first we must define progress on each component of a route in a dynamic route structure. Fix a dynamic route structure (τ^d, l, R) . The basic idea is this. From departure time to the end of the first link, we follow the differential equation for congestion for the first link, and then begin on the second link, and so forth. For notational simplicity, for $i = 1, \dots, l(c)$, define $\tau_i(c)$ to be the time that node $\pi_i(R(c))$ is reached. Evidently, $\tau_1(c) = \tau^d(c)$.

Given a dynamic commuting route structure (τ^d, l, R) , we shall associate with it a function $\delta_{mn}(\tau_m(c), \tau)$ that gives as its value the distance travelled on link mn by commuter c at time τ who begins travel on link mn at time $\tau_m(c)$. In the end, this function will increase in its second argument but decrease in its first argument. Does such a function exist, and is it unique? Fix such a function $\hat{\delta}_{mn}$. To ease notation, compute inductively

$$\tau_{i+1}(c) = \inf\{\tau' > 0 \mid \delta_{\pi_i(R(c))\pi_{i+1}(R(c))}(\tau_i(c), \tau') = \lambda(\pi_i(R(c)), \pi_{i+1}(R(c)))\} \quad (2)$$

We can then compute its *flow at time τ on link mn at distance Δ* , called $\hat{f} : \mathcal{N}^2 \times \mathbb{R}_+^2 \rightarrow \mathbb{R}_+$.¹⁵ It is given by the (possibly discontinuous) partial

¹⁵In terms of notation, f will be a scalar representing an arbitrary value of the flow, whereas \hat{f} is a flow function.

differential equation or *conservation law*:

$$\frac{\partial \hat{f}(m, n, \tau, \Delta)}{\partial t} + \frac{\partial \Phi(m, n, \tau, \Delta)}{\partial \Delta} = 0 \quad (3)$$

where

$$\Phi(m, n, \tau, \Delta) \equiv v \left(\hat{f}(m, n, \tau, \Delta), x(m, n) \right) \cdot \hat{f}(m, n, \tau, \Delta) \quad (4)$$

is defined to be the *flux*. The flux is the volume of commuters passing through a point per unit of time. We abuse notation slightly and sometimes write

$$\Phi_{mn}(f) = v(f, x_{mn}) \cdot f$$

For our example, note that $\Phi_{mn} = \min \{x_{mn}, f\}$.

Equation (3) is actually *the fundamental conservation law of transportation economics* applied to this model. As explained in Bressan (2000, equation 1.2), if we fix an interval of locations on a link, the measure of commuters inside this interval can only change over time from inflows into the interval from the left and outflows from the interval to the right. Another interpretation of equation (3) states that the change with respect to time in commuter density at a given place and time can be found by looking at the change in the flux (commuters per hour) at preceding locations nearby.

Next we compute

$$\frac{\partial \hat{\delta}_{mn}(\tau_m(c), \tau)}{\partial \tau} = v \left(\hat{f}(m, n, \tau, \hat{\delta}_{mn}(\tau_m(c), \tau)), x(m, n) \right) \quad (5)$$

This describes the progress made by commuters on each link of the entire dynamic commuting route structure for any time τ . This is the coupled discontinuous differential equation discussed in the introduction.

Unfortunately, the coupled system defining \hat{f} and $\hat{\delta}$, namely (2), (3), and (5), is technically challenging. The reason is that we cannot restrict τ^d , the function defining the departure strategies of players, beyond assuming that it is a measurable function. Each individual makes a choice, and this is not necessarily coordinated. Discontinuities in departure flows or densities can result in discontinuities in $\partial \hat{\delta}_{mn} / \partial \tau$ that rule out our ability to use standard techniques from the theory of ordinary differential equations as well as the contraction mapping theorem. Instead, we use Bressan (1988).

Even if we can retrieve a well-defined $\hat{\delta}_{mn}$ for each τ^d function, the issue then becomes the fact that there might not exist a Nash equilibrium in pure strategies, since the space of pure strategies is a continuum. Schmeidler (1973)

relies heavily on the fact that the number of pure strategies available to players is finite.

We solve both of the problems at once by simplifying the dynamic model. Fix $\bar{\tau}$ where $T/\bar{\tau}$ is an even integer, and define the departure strategy space to be $\{\bar{\tau}, 3\bar{\tau}, \dots, (T/\bar{\tau} - 1)\bar{\tau}\}$. This makes the strategy space finite. We assume that all the commuters who choose, say, $\bar{\tau}$ will be randomly and uniformly distributed on $(0, 2\bar{\tau})$, those who choose the strategy $3\bar{\tau}$ will be randomly and uniformly distributed on $(2\bar{\tau}, 4\bar{\tau})$, and so forth. The examples in the introduction and that follow fit this framework because they use a uniform distribution of departure times.

We begin by giving the intuition for speed calculations, and then provide a formal proof of existence and uniqueness of the function \hat{f} , from which everything else can be calculated. For example, $\partial \hat{\delta}_{mn} / \partial \tau$ can be calculated from (5). We emphasize that that the description that follows *is provided for intuition about how the system works only*. The formal proof that given departure times and routes, the flows exist and are uniquely determined is technical and does not rely at all on the intuition we provide.

For speed calculations, it is useful to define some concepts. A *threshold* is a location on the network where the speed of commuters is different on the two sides of the threshold at a given time. An important example of a threshold is a node. Of course, a node is a form of a stationary threshold, since it doesn't move over time. Next we will investigate thresholds that move, appear and disappear. An example of a threshold of this type is the boundary between two cohorts, where a *cohort* is defined as a group of commuters with the same route and departure time choices.

Fix a dynamic commuting route structure (τ^d, l, R) . Let $\hat{\tau}^d(c, \tau') = \tau^d(c) + \tau'$, where τ' is a random variable uniformly distributed on $(-\bar{\tau}, \bar{\tau})$, denote the actual departure time of commuter c , that differs from the chosen departure time $\tau^d(c)$ by at most $\bar{\tau}$ as described just above. To reduce the notational burden, we shall generally suppress the second argument (τ') in any function $\hat{\tau}$. Then $\hat{\tau}_1(c) = \hat{\tau}^d(c)$. In general, given $\hat{\tau}_i$, we will define inductively $\hat{\tau}_{i+1}$. Fix any origin node m and destination node $n \neq m$. On each segment mn , define a set of commuters who travel together on a link as:

$$\begin{aligned} \alpha_{mn}(c) &\equiv \{c' \in C \mid \tau^d(c') = \tau^d(c); \text{ for some } i \leq l(c), \\ \pi_1(R(c)) &= \pi_1(R(c')), \dots, \pi_{i-1}(R(c)) = \pi_{i-1}(R(c')); \\ \pi_{i-1}(R(c)) &= \pi_{i-1}(R(c')) = m, \pi_i(R(c)) = \pi_i(R(c')) = n \} \end{aligned}$$

Assume for this informal description that Φ_{mn} is strictly increasing in f . Then the *default speed* for commuter c is given by

$$S_{mn}(c) = v \left(\Phi_{mn}^{-1} (\mu(\alpha_{mn}(c))/2\overline{\tau}), x(m, n) \right)$$

The default speed might be counterfactual, but it is a useful construct. At the default speed, intervals of commuters never overlap with each other. When they never overlap, the time on this link is exactly $\lambda(m, n)/S_{mn}(c)$, so $\widehat{\tau}_{i+1}(c) = \widehat{\tau}_i(c) + \lambda(m, n)/S_{mn}(c)$. Similarly, $\delta_{mn}(\widehat{\tau}_i(c), \tau) = S_{mn}(c) \cdot [\tau - \widehat{\tau}_i(c)]$ where $\pi_i(R(c)) = m$. But there are two other possibilities beyond this first case. The second case is when commuters using different routes blend with each other or separate beginning at a node; this is actually a generalization of the concept of default speed. The third case is if a segment of commuters catches up with another along a link. We consider each of these in turn.

The second case that is possible in the model is when commuters using different routes blend or separate at a node. For the case where they separate, if they are not combined with commuters using other routes, they move at the default speed on the link. But this is just to give intuition. Formally, defining the set of commuters approaching link mn from link $m'm$ at the same time:

$$\begin{aligned} \gamma_{m'mn}(c, \epsilon) \equiv \{ & c' \in C \mid \delta_{m'm}(\widehat{\tau}_{j-1}(c'), \widehat{\tau}_i(c)) \in (\lambda(m', m) - \epsilon, \lambda(m', m)); \\ & \pi_i(R(c)) = m, \pi_{i+1}(R(c)) = n; \\ & \pi_j(R(c')) = m, \pi_{j+1}(R(c')) = n \text{ and } \pi_{j-1}(R(c')) = m' \} \end{aligned}$$

the speed of commuters is given by:

$$S_{mn}^*(c) = v \left(\sum_{m' \neq m} \lim_{\epsilon \rightarrow 0} \frac{\mu(\gamma_{m'mn}(c, \epsilon))}{\epsilon} + \Phi_{mn}^{-1} (\mu(\alpha_{mn}(c))/2\overline{\tau}), x(m, n) \right)$$

Provided that they don't catch up with anyone else, their time on the link is exactly $\lambda(m, n)/S_{mn}^*(c)$, so $\widehat{\tau}_{i+1}(c) = \widehat{\tau}_i(c) + \lambda(m, n)/S_{mn}^*(c)$ whereas $\delta_{mn}(\widehat{\tau}_i(c), \tau) = S_{mn}(c) \cdot [\tau - \widehat{\tau}_i(c)]$ where $\pi_i(R(c)) = m$. This is actually the most general form of the speed and time functions. Notice that since the number of types is finite, the denominator of the right hand side of the last equation actually is almost surely constant for ϵ sufficiently small.

On each segment mn , we say that *commuter c catches up with commuter*

c' on link mn if

$$\begin{aligned}\pi_i(R(c)) &= \pi_j(R(c')) = m, \pi_{i+1}(R(c)) = \pi_{j+1}(R(c')) = n \\ \widehat{\tau}_j(c') &< \widehat{\tau}_i(c) \\ \frac{\lambda(m, n)}{S_{mn}^*(c) - S_{mn}^*(c')} &< \widehat{\tau}_i(c) - \widehat{\tau}_j(c')\end{aligned}$$

The slower commuter, who is unaffected, continues on at the same speed as before the faster one catches them. If commuter c catches up with commuter c' on link mn , define the *catch up*¹⁶ time, for $\pi_i(R(c)) = \pi_j(R(c')) = m$, $\pi_{i+1}(R(c)) = \pi_{j+1}(R(c')) = n$, as $\tau^* = \widehat{\tau}_i(c) + \frac{S_{mn}^*(c') \cdot [\widehat{\tau}_j(c) - \widehat{\tau}_i(c')]}{S_{mn}^*(c) - S_{mn}^*(c')}$. At the first time when a member of one cohort (defined above) catches up with a member of another cohort along a link, a new threshold is created at this time and distance. As it crosses the threshold, the traffic in the faster cohort slows down to the speed of the cohort immediately in front of them by increasing its density at the threshold to match that in the slower cohort. Thus, for all $c'' \in C$ with $\pi_k(R(c'')) = m$, $\pi_{k+1}(R(c'')) = n$, then $\widehat{\tau}_{k+1}(c'') = \tau^* + \frac{\lambda(m, n) - \delta_{mn}(\widehat{\tau}_i(c), \tau^*)}{S_{mn}^*(c')}$ whereas $\delta_{mn}(\widehat{\tau}_k(c''), \tau) = \delta_{mn}(\widehat{\tau}_i(c), \tau^*) + [\tau - \tau^*] \cdot S_{mn}^*(c')$ for all $\tau > \tau^*$ on this link.

The threshold itself moves along the link at speed

$$S_{mn}^*(c') - \frac{S_{mn}^*(c) \cdot \lim_{\epsilon \rightarrow 0} \widehat{f}(m, n, \tau - \epsilon, \delta_{mn}(\widehat{\tau}_k(c''), \tau - \epsilon))}{\lim_{\epsilon \rightarrow 0} \widehat{f}(m, n, \tau - \epsilon, \delta_{mn}(\widehat{\tau}_k(c), \tau - \epsilon))}.$$

We shall remark on this further after a formal statement of the first result.

To prepare for this first result, let us make explicit the assumptions we will use.

Assumption 1: For each fixed x_{mn} , speed $0 < v(f, x_{mn}) < \infty$ is Lipschitz continuous and non-increasing in f .

Assumption 1 means that car speed with no congestion is bounded, speed is a continuous (though not necessarily smooth) function of congestion, and speed does not increase with more cars. As an alternative to assuming that v is Lipschitz, we could directly assume that Φ is Lipschitz, as that is what we use. But since both f and v are bounded (see below after Assumption 2), v Lipschitz implies that Φ is Lipschitz.

Next, we need some preparation for Assumption 2. Eventually, we will need a bound on the total variation of boundary conditions at the start of a

¹⁶Also known as ketchup.

link that is uniform across links. The purpose is to have a compact space that we will use to find a fixed point. A sufficient condition (and necessary for uniform boundedness) is a hierarchy of links that we will specify next. Let the set of links¹⁷ be denoted by:

$$\mathcal{L} \equiv \{(m, n) \in \mathcal{N} \times \mathcal{N} \mid m \neq n\}$$

We postulate a complete preorder on \mathcal{L} denoted by \succeq , with its asymmetric part denoted by \succ . Recall that \mathfrak{R} is the set of all possible routes. Next, we shall restrict routes to $\mathcal{R} \subseteq \mathfrak{R}$.

Assumption 2: Routes r are restricted to:

$$r \in \mathcal{R} \equiv \bigcup_{l=2}^N \{r \in \mathfrak{R}^l \mid \text{For all } i = 2, 3, \dots, l-1, (\pi_i(r), \pi_{i+1}(r)) \succ (\pi_{i-1}(r), \pi_i(r)), \\ \pi_i(r) \neq \pi_j(r) \text{ for } i \neq j\}$$

There are two pieces to this assumption. First, we have restricted route length to N or less. In fact, all that is needed is a finite upper bound on route length. We choose N for simplicity. The assumption that nodes are not repeated along a route makes indexing progress along the route easy. These assumptions are made mainly to keep notation simple.

The second piece is more interesting. Let us begin with the mathematics. The purpose of this assumption is to provide a uniform upper bound on total variation (across time) of boundary or entry conditions for the node at the start of a link. Without this upper bound, we lose both compactness of the space of initial conditions and the ability to solve the differential equation (5). We need compactness for a fixed point theorem, and the ability to solve the differential equation in order to compute travel times and payoffs.

To obtain such an upper bound, we must examine behavior when cohorts merge at a node and travel the next link together. Variation in density in one cohort can be transmitted to the other at the initial node. Thus, total variation can build up. Even if commuters don't travel in circles, the variation that is transmitted can build up along links. So to prevent this, we impose a hierarchy on links.

Turning next to the economics of this assumption, it means that commuters (or packets for the internet) must not be travelling on links that form circles.

¹⁷Also known as genus lynx.

However, travel in opposite directions on links or routes is fine; in fact, this is common for the internet application. For example, if there is a central business district, then one way to satisfy the assumption is to have commuters from each suburb travel towards it during morning rush hour and away from it in the evenings. Circular roads or links forming a circle are also fine, as long as the circle is not completed by overlapping commuters. In the context of the internet, the assumption provides a warning concerning the potential build up of total variation in circles on the internet, even if no set of packets travels in a circle, due to the transmission of total variation across cohorts that travel the same link at the same time.

All of our examples (including directed trees of various sorts used in section 3.2) satisfy this assumption.

Turning next to analysis of the system, there are two immediate, useful consequences of bounding the commuting route length by N . First, the set of routes that are possible for a commuter to choose, henceforth called \mathcal{R} , is finite. Second, we can examine bounds on our endogenous function $\hat{f}(m, n, \tau, \Delta)$. Evidently, $\hat{f}(m, n, \tau, \Delta) \geq 0$. Now consider upper bounds. An upper bound for departure density is $\Phi_{mn}^{-1}\left(\frac{\mu(C)}{2\tau}\right)$. But it is useful to have a uniform bound on density beyond departure density. As we have seen, when one cohort of commuters catches up with a slower cohort ahead of it, this cohort of commuters slows down by building up density so it is the same as that of the slower cohort. Thus, this does not change the upper bound on density. Where density can build up is at nodes, where cohorts can combine. It is important to note also that boundary conditions at the origin of any route are stated in terms of volume (cars per hour) rather than density (cars per mile). Thus, an upper bound on endogenous density is given by the maximal density: $\bar{f} = N \cdot \max_{m,n} \Phi_{mn}^{-1}\left(\frac{\mu(C)}{2\tau}\right)$.

Definition: Let \bar{t} be an upper bound on the time it will take until the last commuter reaches the end of their route:

$$\bar{t} = T + N \cdot \max_{m,n \in \mathcal{N}, m \neq n} \left[\frac{\lambda(m, n)}{v(\bar{f}, x(m, n))} \right]$$

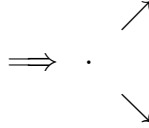
This time will be finite as long as $v > 0$.

At this point, there is an important but technical issue that must be addressed. We shall use Schauder's theorem¹⁸ to show that for any choice of

¹⁸See Smart (1974).

strategies by commuters, namely the choice of route and departure time for each, the flow on each link of commuters in space and time as well as total commuting time are well-defined, namely such a flow exists and is unique. This requires some continuity of commuting times in initial conditions on a link. Moreover, we employ Schmeidler's theorem to prove existence of Nash equilibrium for the dynamic commuting game. One of the requirements of Schmeidler's results is that utility is continuous (in the weak topology on L^1) in the strategy profile of all commuters. For the dynamic model as stated, there is an important type of discontinuity that must be addressed.

The discontinuity is related to moving thresholds. In particular, if a threshold moves backward through a node, a discontinuity in commuting times and payoffs can result. Consider the following example:



Traffic moves from left to right, through a node represented by \cdot . After passing through the node, some traffic heads up and to the right, whereas other traffic heads down and to the right. Suppose that after passing through the node, traffic heading down and to the right flows at high speed, and this traffic volume is steady. Suppose further that a large, slow cohort passes through the node and heads up and to the right, but is followed along the same route by a faster cohort that catches up to the slower one along the upper right link, after the node. Thus, a threshold is formed and the faster cohort slows down to match the speed and density of the slower one. If the volume of this faster cohort is so large that the threshold backs up along the upper right link and through the node to the left link, we claim that a discontinuity in the speed and payoffs of traffic heading down and to the right can occur. The speed of the steady traffic heading down and to the right is reduced to the speed of traffic at the threshold, thus increasing in a discrete manner its density and the time needed to travel the link down and to the right. This can happen despite the fact that much of the traffic on the first link proceeds up and to the right, because the density of traffic using the lower link jumps up when the threshold reaches the left link.

A sufficient (but not necessary) condition to prevent this type of discontinuity would be one that prevents thresholds from moving backwards, whether through a link or not. Thus, we assume:

Assumption 3: $\Phi_{mn}(f)$ is a non-decreasing function of f .

This assumption prevents thresholds from moving backwards, because it says that the volume of consumers moving past a given point will not decrease if density goes up. The direction of movement of thresholds is governed by local volume (cars per unit time), not by local density (cars per unit distance). With this additional assumption (to address the discontinuity), commuting times and payoff functions will be shown to be continuous in strategies.¹⁹

For our examples, note that $\Phi_{mn}(f) = \min \{x_{mn}, f\}$ satisfies this assumption.

Unlike the entire extant literature, we state origin departure boundary conditions (cars entering a link per hour) in terms of volume rather than density (cars per mile). This is a very important distinction. A strategy profile of commuters determines initial volume, namely departures per hour, and not departure density, unless density is completely determined by volume; see Example 4. In general, both departure volume and departure density must be specified in Nash equilibrium.

Definitions: Let

$$\Omega \equiv \left\{ (\hat{\tau}, \tau) \in [0, \bar{t}]^2 \mid \tau \geq \hat{\tau} \right\} \quad (6)$$

$$\mathcal{D}_{mn} \equiv \left\{ \delta_{mn} : \Omega \rightarrow [0, \lambda(m, n)] \text{ measurable} \mid \text{for } (\hat{\tau}, \tau), (\hat{\tau}', \tau') \in \Omega: \delta_{mn}(\hat{\tau}, \hat{\tau}) = 0, \right. \\ \left. \left| \delta_{mn}(\hat{\tau}, \tau) - \delta_{mn}(\hat{\tau}', \tau') \right| \leq v(0) \cdot (|\hat{\tau} - \hat{\tau}'| + |\tau - \tau'|) \right\}$$

$$\mathcal{D} \equiv \prod_{m, n=1, m \neq n}^N \mathcal{D}_{mn}$$

We use square block metric for the Lipschitz condition as a matter of convenience.

The following definition comes from Strub and Bayen (2006), adapted to our context. Interpretations immediately follow the definitions. For further discussion, see also Bressan (2000).

Definition: A collection of measurable functions $\left\{ \hat{f}(m, n, \cdot, \cdot), \hat{\delta}_{mn} \right\}_{m, n=1, m \neq n}^N$, where $\hat{f}(m, n, \cdot, \cdot) : [0, \bar{t}] \times [0, \lambda(m, n)] \rightarrow [0, \bar{f}]$ and $\hat{\delta}_{mn} \in \mathcal{D}_{mn}$, is called

¹⁹We conjecture that it would be possible to allow such discontinuities and omit Assumption 3 using the following technique. Strub and Bayen (2006) allow a boundary condition at the end of a link as well as the beginning of a link, and this can be combined with Khan's (1989) generalization of Schmeidler (1973) to upper semicontinuous payoff functions to obtain existence of Nash equilibrium. However, as Strub and Bayen (2006, section 2) note, boundary conditions on both ends of a link can cause inconsistencies in the flow on a link, and the possibility of no solution. That is why they use a *weak* formulation of boundary conditions that allows violation of such boundary conditions under certain circumstances.

a **solution** to the conservation law (3) with initial and boundary conditions if, for every m and n ($m \neq n$), for every $k \in \mathbb{R}$, for every C^1 function $\varphi_{mn} : [0, \bar{t}] \rightarrow \mathbb{R}_+$, for every C^1 function $\psi_{mn} : [0, \bar{t}] \times [0, \lambda(m, n)] \rightarrow \mathbb{R}_+$, the following hold:

$$0 \leq \int_0^{\lambda(m, n)} \int_0^{\bar{t}} \left| \widehat{f}(m, n, \tau', \delta') - k \right| \cdot \frac{\partial \psi_{mn}(\tau', \delta')}{\partial \tau} \quad (7)$$

$$+ \text{sign} \left(\widehat{f}(m, n, \tau', \delta') - k \right) \cdot \left[\Phi_{mn} \left(\widehat{f}(m, n, \tau', \delta') \right) - \Phi_{mn}(k) \right] \cdot \frac{\partial \psi_{mn}(\tau', \delta')}{\partial \delta} d\tau' d\delta'$$

and there exist $2(N-1)^2$ sets of Lebesgue measure zero: $E_{mn}^0 \subseteq [0, \lambda(m, n)]$, $E_{mn}^L \subseteq [0, \bar{t}]$, such that for all $m, n = 1, \dots, N$, $m \neq n$,

$$\lim_{\tau \rightarrow 0, \tau \notin E_{mn}^0} \int_0^{\lambda(m, n)} \left| \widehat{f}(m, n, \tau, \delta') \right| d\delta' = 0$$

$$\lim_{\delta \rightarrow 0, \delta \notin E_{mn}^L} \int_0^{\bar{t}} L_{mn} \left(\widehat{f}(m, n, \tau', \delta), \rho_{mn}(\tau') \right) \varphi(\tau') d\tau' = 0$$

where

$$L_{mn}(a, b) \equiv \sup_{\kappa \in I(a, b)} (\text{sign}(a - b) \cdot [\Phi_{mn}(a) - \Phi_{mn}(\kappa)])$$

$$I(a, b) \equiv [\inf(a, b), \sup(a, b)]$$

$$\widehat{\delta}_{mn}(\widehat{\tau}, \tau) = \int_{\widehat{\tau}}^{\tau} v \left(f(m, n, \bar{\tau}, \widehat{\delta}_{mn}(\widehat{\tau}, \bar{\tau})), x_{mn} \right) d\bar{\tau}$$

$$\bar{\mu}_r^i(\tau) \equiv \mu \left(\{c \in C \mid \exists c' \in C \text{ with } R(c') = R(c) = r, \tau_i(c') = \tau, \tau^d(c') = \tau^d(c)\} \right)$$

$$\rho_{mn}(\tau) \equiv \Phi_{mn}^{-1} \left(\frac{\mu(\{c \in C \mid \pi_1(R(c)) = m, \pi_2(R(c)) = n, |\tau^d(c) - \tau| < \bar{\tau}\})}{2\bar{\tau}} \right) + \quad (8)$$

$$\sum_{m'=1, m' \neq m}^N \frac{\sum_{\{r \in \mathcal{R} \mid \text{for some } i \pi_{i-1}(r)=m', \pi_i(r)=m, \pi_{i+1}(r)=n\}} \bar{\mu}_r^i(\tau)}{\sum_{\{r \in \mathcal{R} \mid \text{for some } i \pi_{i-1}(r)=m', \pi_i(r)=m\}} \bar{\mu}_r^i(\tau)} \cdot \widehat{f}(m', m, \tau, \lambda(m', m))$$

Remark 1: The crucial but subtle connection between the functions \widehat{f} and $\widehat{\delta}$ is through the last boundary condition and definition (2). The last condition gives entry into a link by those just departing from their origin node and those continuing their travel through the node from other links.

Remark 2: What we call a solution is actually a refinement of other solution concepts used in the literature that are more obviously related to (3). The least restrictive of these is the concept of distributional solution, followed by the more restrictive weak solution. The (yet more restrictive) solution concept we

use is generally called an entropy weak solution in the literature. Motivation for using this solution is that although we have existence theorems for all of these solution concepts, uniqueness holds only for the entropy weak solution. There is also intuition for the refinement in terms of stability, usually called admissibility conditions, in the literature we have cited.

Remark 3: It is important to provide at least a heuristic explanation, part of the folklore in the literature, about why this represents a solution to the partial differential equation or conservation law (3), since there is no obvious connection between the partial differential equation and what we call a solution.²⁰ Suppose that ψ can be chosen so that $\frac{\partial \psi_{mn}}{\partial \tau}$ is close to an indicator function for some set in $[0, \bar{t}] \times [0, \lambda(m, n)]$ and $\frac{\partial \psi_{mn}}{\partial \delta}$ is close to an indicator function for that same set multiplied by $\frac{1}{v(\hat{f}(m, n, \tau, \delta), x_{mn})}$, so that we can focus on the integrand in inequality (7). If we can choose another function so that these derivatives are close to -1 multiplied by these functions,²¹ then inequality (7) implies:

$$\left| \hat{f}(m, n, \tau, \delta) - k \right| + \text{sign} \left(\hat{f}(m, n, \tau, \delta) - k \right) \cdot \left[\Phi_{mn} \left(\hat{f}(m, n, \tau, \delta) \right) - \Phi_{mn}(k) \right] \cdot \frac{1}{v \left(\hat{f}(m, n, \tau, \delta), x_{mn} \right)} = 0$$

Dividing by $\text{sign} \left(\hat{f}(m, n, \tau, \delta) - k \right)$, we obtain

$$\left(\hat{f}(m, n, \tau, \delta) - k \right) + \left[\Phi_{mn} \left(\hat{f}(m, n, \tau, \delta) \right) - \Phi_{mn}(k) \right] \cdot \frac{1}{v \left(\hat{f}(m, n, \tau, \delta), x_{mn} \right)} = 0$$

Now choose $k_h = \hat{f}(m, n, \tau - \frac{1}{h}, \delta)$ for $h = 1, 2, 3, \dots$. Then dividing by $\frac{1}{h}$ and taking limits as $h \rightarrow \infty$ yields

$$\frac{\partial \hat{f}(m, n, \tau, \delta)}{\partial \tau} + \Phi'_{mn} \left(\hat{f}(m, n, \tau, \delta) \right) \cdot \frac{1}{v \left(\hat{f}(m, n, \tau, \delta), x_{mn} \right)} = 0$$

This expression is the same as (3).

Theorem 1: Suppose that v satisfies Assumption 1 and that feasible routes are restricted to satisfy Assumption 2. Suppose further that flux Φ_{mn} is strictly

²⁰Evidently, this is one of the barriers to entering this literature.

²¹Notice that these restrictions are on the derivatives of ψ_{mn} rather than on ψ_{mn} itself, so it is possible to make the derivatives negative while satisfying the non-negativity constraint on ψ .

increasing in density f . Then to each dynamic commuting route structure (τ^d, l, R) , there corresponds a unique solution $\left\{ \widehat{f}(m, n, \cdot, \cdot), \widehat{\delta}_{mn} \right\}_{m,n=1, m \neq n}^N$.

Remark 4: The case where Φ_{mn} is weakly increasing, as in the examples, will be dealt with when existence of equilibrium is considered. For technical reasons, it is easiest to consider this case as a limit of the cases where Φ_{mn} is strictly increasing.

Remark 5: One issue concerning our system is how we define a solution. Our system in f is generally rather discontinuous, so it requires special treatment. There are alternatives to the technique we use, which we consider to be the most straightforward given our framework. One such alternative is to assume that the flux function Φ_{mn} is smooth and either strictly convex or strictly concave. The conservation law is then called *strictly hyperbolic*; see Bressan (2000), particularly section 10.2. We can then define a *Filippov solution* (Filippov, 1973) to this problem, that was introduced into economics by Ito (1979).²² Colombo and Marson (2003) and particularly Marson (2004) can be applied to obtain existence and uniqueness of a solution.²³ However, we do not place further restrictions on the flux.

The proof of Theorem 1 can be found in the Appendix. Formally, we prove that for given departure times and route choices the system behavior given by $\left\{ \left(\widehat{f}(m, n, \cdot, \cdot), \widehat{\delta}_{mn}(\cdot, \cdot) \right) \right\}_{m,n=1}^N$ exists and is uniquely defined. To accomplish this, we apply Schauder's theorem in a slightly unorthodox manner to the set of boundary conditions for each node, where the boundary conditions lie in the space of functions of bounded variation with respect to time.

Next we examine existence of Nash equilibrium in pure strategies in our dynamic model.

The *time cost of a dynamic commuting structure* (τ^d, l, R) for commuter c is $\int_{-\bar{\tau}}^{\bar{\tau}} \frac{\widehat{\tau}_{l(c)}(c, \tau) - (\tau^d(c) + \tau)}{2\bar{\tau} \cdot \mu(\{c' \in C \mid \pi_1(R(c)) = \pi_1(R(c')), \pi_2(R(c)) = \pi_2(R(c')), \tau^d(c) = \tau^d(c')\})} d\tau$. In essence, this is the expected time cost taken over all commuters using the same pure departure (time and first road) strategy. These are the commuters that determine departure volume of a cohort with commuter c on the link.

²²Formally speaking, we could introduce the general definition of a Filippov solution and then show that there exists one with finite total variation, but here we follow Colombo and Marson (2003) and Marson (2004) who skip this step because this fact is already well-known.

²³In fact, Strub and Bayen (2006) use a strictly concave flux function in their application in section 5 to the I-210 in Los Angeles. Thus, they could have used a Filippov solution instead of a weak entropy solution.

We fix an arrival time at $\tau^A \in [0, \infty]$. Next we introduce the arrival penalty function $P : \mathbb{R}_+ \rightarrow \mathbb{R}_+$. To give intuition, think of $\tau = \tau_{l(c)}(c)$. The arrival penalty is given by

$$P(\tau) \geq 0 \text{ where } P(\tau^A) = 0$$

For example, in the introduction we required that:

$$\text{Almost surely for } c \in C, \hat{\tau}_{l(c)}(c) \leq \tau^A$$

Thus, $P(\tau) = 0$ for $\tau \leq \tau^A$ whereas $P(\tau) = \infty$ for $\tau > \tau^A$. It is actually more common in the literature to use an asymmetric linear penalty function; see Arnott et al (1993). We can allow further generalization, for example heterogeneous required arrival times τ^A , but at the cost of messier notation. We note that in the framework with a finite number of departure times, this is actually the *expected* penalty for the given choice of strategy, since commuters are randomly assigned over a small departure time interval.

The individual payoff function for the dynamic model is thus:

$$\begin{aligned} u(c; \tau^d, l, R) \equiv & \\ & - \int_{-\bar{\tau}}^{\bar{\tau}} \frac{\hat{\tau}_{l(c)}(c, \tau') - (\tau^d(c) + \tau') + P(\hat{\tau}_{l(c)}(c, \tau'))}{2\bar{\tau} \cdot \mu(\{c' \in C \mid \pi_1(R(c)) = \pi_1(R(c')), \pi_2(R(c)) = \pi_2(R(c')), \tau^d(c) = \tau^d(c')\})} d\tau' \end{aligned} \quad (9)$$

The *utilitarian welfare function for the dynamic model* is

$$\begin{aligned} U(\tau^d, l, R) = & \\ & - \int_C \int_{-\bar{\tau}}^{\bar{\tau}} \frac{[\hat{\tau}_{l(c)}(c, \tau') - (\tau^d(c) + \tau') + P(\hat{\tau}_{l(c)}(c, \tau'))]}{2\bar{\tau} \cdot \mu(\{c' \in C \mid \pi_1(R(c)) = \pi_1(R(c')), \pi_2(R(c)) = \pi_2(R(c')), \tau^d(c) = \tau^d(c')\})} d\tau' d\mu(c) \end{aligned}$$

A *Nash equilibrium in pure strategies of the dynamic model* is a dynamic commuting structure (τ^d, l, R) such that almost surely for $c \in C$, there is no route r of length $v \leq N$ and departure time $\tau^{d'}$ for commuter c such that, computing arrival times $\hat{\tau}'$ as in Theorem 1 for the new route and departure time,

$$\begin{aligned} & \int_{-\bar{\tau}}^{\bar{\tau}} \frac{\hat{\tau}_{l(c)}(c, \tau') - (\tau^d(c) + \tau') + P(\hat{\tau}_{l(c)}(c, \tau'))}{2\bar{\tau} \cdot \mu(\{c' \in C \mid \pi_1(R(c)) = \pi_1(R(c')), \pi_2(R(c)) = \pi_2(R(c')), \tau^d(c) = \tau^d(c')\})} d\tau' > \\ & \int_{-\bar{\tau}}^{\bar{\tau}} \frac{\hat{\tau}'_v(c, \tau') - (\tau^{d'} + \tau') + P(\hat{\tau}'_v(c, \tau'))}{2\bar{\tau} \cdot \mu(\{c' \in C \mid \pi_1(R(c)) = \pi_1(R(c')), \pi_2(R(c)) = \pi_2(R(c')), \tau^d(c) = \tau^d(c')\})} d\tau' \end{aligned}$$

We note that due to the congestion externality, the Nash equilibria are unlikely to be Pareto (or utilitarian) optimal. Example 2 below will make this precise.

Next, in Theorem 2, we shall prove existence of Nash equilibrium in pure strategies for our model with discrete and finite departure times by applying Schmeidler (1973, Theorems 1 and 2). For the model with a continuum of departure time strategies, we can only obtain existence of ϵ -equilibrium in pure strategies. It is also worth noting that since Φ_{mn} will not be required to be strictly increasing, we must modify (8) to:

$$\rho_{mn}(\tau) \in \Phi_{mn}^{-1} \left(\frac{\mu(\{c \in C \mid \pi_1(R(c)) = m, \pi_2(R(c)) = n, |\tau^d(c) - \tau| < \bar{\tau}\})}{2\bar{\tau}} \right) + \sum_{m'=1, m' \neq m}^N \frac{\sum_{\{r \in \mathcal{R} \mid \text{for some } i \pi_{i-1}(r)=m', \pi_i(r)=m, \pi_{i+1}(r)=n\}} \bar{\mu}_r^i(\tau)}{\sum_{\{r \in \mathcal{R} \mid \text{for some } i \pi_{i-1}(r)=m', \pi_i(r)=m\}} \bar{\mu}_r^i(\tau)} \cdot \hat{f}(m', m, \tau, \lambda(m', m))$$

This adds another layer of indeterminacy to Nash equilibrium in the case where

Φ_{mn} is not strictly increasing. Consider, for example, $\nu(f, x_{mn}) = \frac{1}{f}$, and thus $\Phi_{mn}(f) = 1$. Then any departure density with respect to distance can be made consistent with the initial volume conditions at the origin node.

Theorem 2: Under Assumptions 1-3, if the penalty function P is continuous,²⁴ there exists a Nash equilibrium in pure strategies.

One can prove that a utilitarian optimum exists for the discrete departure time model. Instead of looking at a continuum of individual strategies, give the social planner the control variables that are the measure of commuters using each route at each departure time. The control vector is finite-dimensional. Assume, to begin, that Φ_{mn} is strictly increasing in f . With assumptions 1-2 and Theorem 1, flows and utility levels are well-defined for each departure and route strategy profile. In the proof of Theorem 2, found in Appendix 2, it is shown that destination arrival times are continuous in the departure and route strategy profile. Thus, the utilitarian objective is continuous as a function of the measure of commuters using each route and departure time, so an optimum exists.

Consider next the case where Φ_{mn} is non-decreasing in f . As usual, take a sequence of initial conditions converging to the supremum. These initial conditions are in terms of volumes and routes, but there exists associated departure densities (per mile instead of per minute) associated with these volumes such that the supremum is approached. In the proof of Theorem 1, the only use made of Φ_{mn} strictly increasing in f is to prove that \hat{f} is unique, so

²⁴Although some of our examples, such as the one in the introduction, feature a discontinuous P , a nearby continuous P with sufficiently steep slope just after the arrival time would work just as well, but would distract from the point of the example.

there is an associated sequence of flows such that the optimum is approached. Following the remainder of the proof of Theorem 2 (that proves continuity of the objective in the strategy profile), the optimum will be achieved in the limit.

Example 2: What does Nash Equilibrium look like in the case of a linear penalty function? This is important for applications, as much of the literature uses such a specification. It is actually quite interesting. Suppose that

$$P(\tau) = \begin{cases} \eta(\tau^A - \tau) & \text{if } \tau^A \geq \tau \\ \psi(\tau - \tau^A) & \text{if } \tau \geq \tau^A \end{cases}$$

where $\eta, \psi > 0$. To fix ideas, we consider the example from the introduction, with one link and two nodes, modified for this penalty function. Capacity of the link is $x = 1$, whereas travel time on the uncongested link is 1. At a Nash equilibrium, utility must be equalized across commuters, for otherwise everyone will imitate the happiest ones only. Fortunately for urban economists, this is a familiar condition. There is mass 2 of identical commuters. Consider an example with 2 departure times, $\frac{1}{2}$ and $\frac{3}{2}$. Those who choose departure time $\frac{1}{2}$ actually leave at a random time distributed uniformly between 0 and 1, whereas those who choose departure time $\frac{3}{2}$ actually leave at a random time distributed uniformly between 1 and 2. Let $\tau^A = \frac{7}{2}$ and $\eta = \psi \leq \frac{1}{2}$. It will turn out that in Nash equilibrium, the commuters who choose departure time $\frac{1}{2}$ travel at the speed limit, whereas the commuters who leave at time $\frac{3}{2}$ travel slower and arrive later. Suppose the (endogenous) measure of commuters who choose departure time $\frac{1}{2}$ is called w , whereas the (endogenous) measure of commuters who leave at time $\frac{3}{2}$ is called w' , where $w + w' = 2$. For those who choose departure time $\frac{1}{2}$, their travel time is 1 whereas their *expected* early arrival penalty is 2η . For those departing at time $\frac{3}{2}$, their travel time is w' whereas their *expected* early arrival penalty is $\eta \cdot (\frac{7}{2} - (w' + \frac{3}{2}))$. Setting these negative utilities equal to each other, we obtain $w' = \frac{1}{1-\eta}$, and thus $w = \frac{1-2\eta}{1-\eta}$. Notice that, similar to Example 1, we can create a Pareto improvement by making more agents choose departure time $\frac{1}{2}$. This disrupts the equal utility condition.

3 Applications

3.1 Can the Static Equilibrium be Supported by a Dynamic Equilibrium?²⁵

Here we ask the following question. Given identical exogenous data for the static and dynamic commuting games and finding Nash equilibrium, are the flows in the static and dynamic models the same? In other words, is the static model a reduced form of the dynamic model? This is important for addressing the issue of whether the static model makes sense. For if the answer to this question is negative, then there should be no interest in the static model, since its equilibrium behavior is different from the analogous dynamic model, and the real world is dynamic.

For simplicity, we return now to the examples used in many of the previous sections, namely where there is no penalty for early arrival and an infinite penalty for late arrival. One could imagine that the static model represents some sort of steady state of the dynamic model, where commuters are introduced at constant flows at all the nodes, and the flows in the links are constant over time. But there are two problems with this idea. First, with a fixed arrival time (say 9 AM), a steady state does not make sense. The time profile of equilibrium departures will generally not be constant over time, since everyone must get to work by the arrival time. Even if arrival time varied by commuter, one would not expect to see a steady state necessarily attained. Second, the two alternative concepts for consistency of the two models we introduce next are *weaker* than asking that a steady state of the dynamic model look like a static equilibrium. In other words, if a steady state of the dynamic model looked like the static model, then the conditions would be satisfied. But they are not.

One could ask whether average flows (over time and space or distance on a link) in the dynamic model are equilibrium flows of the static model. Given identical exogenous data for the static and dynamic games and finding

²⁵The ideas in this subsection owe much to Anas (2007) and to discussions with Alex Anas.

equilibrium, does the following condition on flows hold?²⁶

$$f(m, n) = \frac{\int_0^{\lambda(m, n)} \int_0^{\tau^A} \widehat{f}(m, n, \tau, \Delta) d\tau d\Delta}{\lambda(m, n) \cdot \tau^A} \text{ for all } m, n = 1, 2, \dots, N$$

But this disguises the following issue. In the dynamic model, flows could be high for a time and then zero. The average over the link and over time would be in between, but there would be no actual time and distance on the link where the average was actually attained. So it is logical to ask whether there is a time τ , and a distance on every link $\Delta(m, n)$, such that the flows from the static model are attained by the dynamic model:²⁷

$$f(m, n) = \widehat{f}(m, n, \tau, \Delta(m, n)) \text{ for all } m, n = 1, 2, \dots, N$$

To answer all of these questions in the negative, one only need go back to the simple example with two nodes and one link given in the introduction. There the uncongested commuting pattern Nash equilibrium is not present for the static model, though the congested one is. But if we want to say something more, for example that there is *no* equilibrium of the dynamic model that replicates the behavior of the static model, then we must become slightly more sophisticated.

Example 3: We set up a network with 3 identical links in series, each one with the structure of the simple example in the introduction (equivalently, one could use 2 nodes and 1 link with the travel time multiplied by 3). Then if we set the arrival time at $\frac{1}{x} + 3$ (where $x < 1$), the congested commuting pattern violates the arrival time for the last commuters, the commuters departing at time 1 (they arrive at $\frac{3}{x} + 1$), and the uncongested commuting pattern remains as the only equilibrium of the dynamic model. It violates all of the conditions above, as there is no uncongested commuting pattern for the static model. In fact, even if we only pay attention to distances on links where there are commuters, their density is $x < 1$, never to be found in an equilibrium of the static model. In summary, for this example, the only equilibrium of the static model is the congested commuting pattern, whereas the only equilibrium of the

²⁶In a steady state of the dynamic model, this condition would be satisfied because the flow on each link would be constant, independent of time, and thus be equal to the average flow.

²⁷In a steady state of the dynamic model, this condition would be satisfied because the flow on each link would be constant, independent of time, so it would be satisfied for every time.

dynamic model is the uncongested commuting pattern. Thus, the equilibrium sets of the two models are unrelated.²⁸

Verhoef (1999) studies a similar problem in a very different class of models, and concludes (p. 365) that, “For static models of peak demand, it was argued that for such models to be dynamically consistent, rather heroic assumptions on the pattern of scheduling costs have to be made.”

3.2 Welfare Properties of Nash Equilibrium

Equilibrium selection is an important issue in one shot congestion games with Pigouvian congestion taxes. Under such taxes, there can be multiple Nash equilibria, only some of which are efficient.²⁹ As remarked in the introduction, Sandholm (2007) shows that with a finite number of commuters, an evolutionary process, and Pigouvian taxes, the outcome will be efficient. A major limitation of this work is the assumption of a common utility function with idiosyncratic perturbations, which seems to rule out heterogeneous origins and destinations.

Although that approach is clearly interesting, we take a completely different approach here, motivated by our examples. A major advantage of our approach is that we can compare non-trivial commutes (home to work) with their reverses (work to home), to our knowledge absent in the literature. As we wish to focus on departure times rather than routes in the dynamic model, we impose the following restriction:

Definitions: A **outbound tree network** is a set of route lengths and routes \mathcal{R} such that for any $(l, r), (l', r') \in \mathcal{R}$, there do not exist $1 < i \leq l$ and $1 < i' \leq l'$ with $\pi_{i-1}(r) \neq \pi_{i'-1}(r')$ and $\pi_i(r) = \pi_{i'}(r')$. An **inbound tree network** is a set of route lengths and routes \mathcal{R} such that for any $(l, r), (l', r') \in \mathcal{R}$, there do not exist $1 \leq i < l$ and $1 \leq i' < l'$ with $\pi_i(r) = \pi_{i'}(r')$ and $\pi_{i+1}(r) \neq \pi_{i'+1}(r')$.

In terms of commuting, an inbound tree network might be a reasonable model of commuting from home to work, whereas an outbound tree network

²⁸Without an arrival time, it’s easy to argue that neither the static nor the dynamic model is a reasonable model of the morning commute.

²⁹We do not provide an example here, both because they are available in the literature (for more macro models) and because, as will be apparent from Theorem 3, examples in our framework with non-constant (or non-zero) Pigouvian taxes will have relatively complicated route structures. For instance, a one link example won’t work.

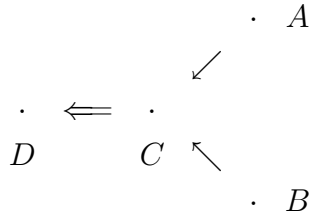
might be a reasonable model of commuting from work to home. In terms of electronic networks, this might not be a good model of the internet, but tree structures are often used in local area networks. The property of interest for an outbound tree network is preventing mergers of routes at nodes where traffic continues together along the next link.

You might wonder why we are introducing a restrictive condition like this. It is one way to sort out the efficiency properties of Nash equilibrium in our dynamic model. What is perhaps strange but interesting is that on a two way network, commuting to work may be inefficient, whereas commuting to home might be efficient. In other words, reversing the commute on a directed network can change the efficiency properties of Nash equilibrium.

We begin our discussion with the static model. With an inbound or outbound tree network, there are no choices to be made, so rather trivially there is an efficient Nash equilibrium. We conjecture that this result can be extended to more general structures than trees³⁰, but that would distract from the main point of this section.

Next we turn to the dynamic model. In particular, we wish to examine the similarities and differences between commuting from home to work and commuting from work to home. Since networks are arbitrary in our general framework, we focus on trees, and begin our analysis with an example. Most of the intuition can be gleaned from this example. What is important for our purposes is asymmetry.

Example 4: First, consider the commute from two home locations A and B on the right to a common work location D on the left, via a merge at node C :



Suppose that rush hour is from time 0 to time 1 with two possible departure times: $1/4$ and $3/4$, where commuters choosing the first departure time are uniformly distributed over actual departure times $[0, 1/2]$ and commuters choosing the second departure time are uniformly distributed over actual departure times $[1/2, 1]$. The length of each link is 1. Speed on links between nodes A and C , as well as between nodes B and D , is given by $\min\left\{\frac{1}{f}, 1\right\}$.

³⁰E.g. shrubberies, with leaves at the ends of branches.

Speed on the link between nodes C and D is given by $\min \left\{ \frac{2}{f}, 2 \right\}$. There is no arrival time penalty; it's not very natural when comparing a commute and its reverse, though the examples and theorem can likely be extended in this direction. There is measure 1 commuters travelling from node A to node D , whereas there is measure 1 travelling from node B to node D .

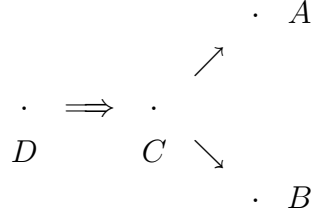
It is obvious that many departure *densities* will be consistent with the departure *volume*, and that departure volume must be 1. Thus, strategy profiles will focus on departure density rather than volume.

Let's first examine Nash equilibrium. Consider the symmetric strategy profile where measure $1/2$ commuters at home A choose departure time $1/4$ and thus are uniformly distributed with flux volume 1 over $[0, 1/2]$, whereas measure $1/2$ commuters at home A choose departure time $3/4$ and thus are uniformly distributed with flux volume 1 over $[1/2, 1]$. As always, we must be careful about how volume (cars per hour) translates into density (cars per mile), particularly at origins. In this case, we set density $f = 2$, so speed is $1/2$. Similarly, commuters at home B are split: measure $1/2$ choose departure time $1/4$ and are uniformly distributed over $[0, 1/2]$ with flux volume 1, whereas measure $1/2$ choose departure time $3/4$ and are uniformly distributed over $[1/2, 1]$ with flux volume 1. Each commuter experiences a total travel time of 4: travel time is 2 on the initial link, and 2 on the link between node C and node D .

Next consider the following slightly asymmetric strategy profile that will not be a Nash equilibrium. The departure *volumes* are the same as for the Nash equilibrium profile, but the departure *densities* will be different. The density schedule for commuters who have homes at A remain the same as above. Commuters who live at B will have the following departure schedule. For those who depart at time $1/4$, the departure density is 1.5, so initial speed is $\frac{1}{1.5}$. For those who depart at time $3/4$, the departure density is 2.5. The first cohort to arrive at node C will be those from homes at node B who depart in $[0, 1/2]$, and who arrive at node C at times in the interval $[1.5, 2]$. Next are the commuters from homes at node A who depart in $[0, 1/2]$, and who arrive at node C at times in the interval $[2, 2.5]$. The next commuters to arrive are the commuters from node A who depart in the interval $[1/2, 1]$ and who arrive at node C at times in the interval $[2.5, 3]$. Finally, the cohort of commuters who depart node B in $[1/2, 1]$ arrive at node C at times in the interval $[3, 3.5]$. Notice that the overlap in arrival times at node C is of measure zero. Therefore, in the end, travel time for all commuters who live at

node A is 3, whereas travel time for the first cohort from node B is $\frac{3}{2} + \frac{4}{3} = \frac{17}{6}$, but for the second cohort is $\frac{5}{2} + \frac{4}{5} = \frac{33}{10}$. Clearly, this strategy profile Pareto dominates the Nash equilibrium strategy profile, but is not a Nash equilibrium itself.

Next we reverse the commute. The structure of permissible departure times and speed as a function of density are the same. The new diagram is as follows:



A Nash equilibrium and utilitarian optimal strategy profile has measure $1/2$ of each type departing work at each of the two departure times. The departure density is 2, whereas initial speed is 1. The total travel time of each commuter is 2.

Theorem 3: With an outward bound tree network, assuming $\frac{\partial v}{\partial f} < 0$ and $P = 0$, there is a Pareto optimal³¹ strategy profile that is also a Nash equilibrium. Thus, there exists an efficient Nash equilibrium.

Thus, under these additional assumptions, efficiency can be achieved not through taxes, but by equilibrium selection. Prisoners' dilemma problems are ruled out by the structure of the game, specifically these additional assumptions. Example 4 is an example of an inward bound tree network that has no efficient Nash equilibrium, so an analog of Theorem 3 for an inward bound tree network is impossible.

Since the simple example with two nodes and one link from the introduction is trivially an outward bound tree network, it cannot be true that all Nash equilibria are efficient.

4 Conclusions

We have asked and answered several questions about commuting using two models, one static and one dynamic. For each model, we have shown that a Nash equilibrium in pure strategies exists for the one shot game, that a Pareto

³¹Although we have not defined it formally, Pareto optimum is the usual concept in our context of a continuum of agents.

optimum exists, and that Nash equilibrium is generally not Pareto optimal. Beyond that, we have shown that all Nash equilibria of the static model can look very different from any Nash equilibrium of the dynamic model. Since the static model features behavior unlike the dynamic one, we reject the former as a reduced form of the latter and stick with the dynamic model. Finally, we have examined the welfare properties of Nash equilibrium in the particular case of a tree network, and found that equilibrium might not be efficient for the morning commute, but there always is an efficient Nash equilibrium for the evening commute. Thus, congestion pricing is more important for the morning commute, whereas equilibrium selection is more important for the evening commute. Further effort should be devoted to the welfare properties of Nash equilibrium on specific directed networks. In sum, what we have shown is that a model of congestion using microfounded behavior has very different properties from the reduced form models used in the literature.

Our commuting model can be reinterpreted as a model of internet congestion. In this context, local area networks often have a tree structure, so for example the results on efficiency of Nash equilibrium and the consequences for congestion tolls can be interpreted in this framework.

For simple examples, the Nash equilibria of our model can be solved analytically. For more complex examples, the proof of Theorem 1 indicates that a numerical solution technique involves nesting the solution of a discontinuous system of differential equations inside a fixed point solution algorithm.

In a companion paper to this research, Berliant (2012) examines the set of Nash equilibria in the infinitely repeated versions of both the static and dynamic commuting games, and the folk theorem is used to obtain these large sets. There we present some preliminary evidence from the shutdown of an expressway in St. Louis that commuters do not always play one shot Nash equilibrium. We also discuss the application of the anti-folk theorem to our specific game, namely conditions under which the Nash equilibria of the infinitely repeated game are the Nash equilibria of the one shot game.

Our model could be extended to allow elastic demand for travel to or from work. The extension of the model to allow land markets and endogenous choice of household residence and job location would also be interesting.

The dynamic model should be applied to real world commuting. Since it can accommodate an arbitrary (exogenous) route structure, it has both positive and normative content, especially regarding Pareto improvements. For example, it can be used to perform cost benefit analysis with respect to chang-

ing infrastructure and mass transit.

References

- [1] Anas, A., 2007. “The Trips-to-Flows Riddle in Static Traffic Equilibrium: How to drive a BMW ?” Unpublished manuscript.
- [2] Arnott, R., A. de Palma and R. Lindsey, 1993. “A Structural Model of Peak-Period Congestion: A Traffic Bottleneck with Elastic Demand.” *American Economic Review* 83, 161-179.
- [3] Beckmann, M., C.B. McGuire and C.B. Winsten, 1956. *Studies in the Economics of Transportation*. Yale University Press: New Haven.
- [4] Berliant, M., 2012. “Daily Commuting.” Working Paper.
- [5] Bressan, A., 1988. “Unique Solutions for a Class of Discontinuous Differential Equations.” *Proceedings of the American Mathematical Society* 104, 772-778.
- [6] Bressan, A., 2000. *Hyperbolic Systems of Conservation Laws: The One-Dimensional Cauchy Problem*. Oxford University Press: Oxford.
- [7] Bressan, A. and W. Shen, 1998. “Uniqueness for Discontinuous ODE and Conservation Laws.” *Nonlinear Analysis* 34, 637-652.
- [8] Colombo, R.M. and A. Marson, 2003. “A Hölder Continuous ODE Related To Traffic Flow.” *Proceedings of the Royal Society of Edinburgh* 133A, 759-772.
- [9] Daganzo, C.F., 2008. *Fundamentals of Transportation and Traffic Operations*. Emerald Group Publishing: Bingley, UK.
- [10] de Palma, A. and M. Fosgerau, 2011. “Dynamic Traffic Modeling.” In *A Handbook of Transport Economics*, edited by A. de Palma, R. Lindsey, E. Quinet and R. Vickerman. Edward Elgar: Cheltenham, UK. Ch. 9, pp. 188-212.
- [11] Filippov, A.F., 1973. *Differential Equations with Discontinuous Righthand Sides*. Kluwer: Dordrecht, Netherlands.

- [12] Garavello, M. and B. Piccoli, 2006. *Traffic Flow on Networks: Conservation Laws Model*. American Institute of Mathematical Sciences: Springfield, MO.
- [13] Han, K., T.L. Friesz and T. Yao, 2013. “Existence of Simultaneous Route and Departure Choice Dynamic User Equilibrium.” *Transportation Research Part B* 53, 17-30.
- [14] Han, K., B. Piccoli and T.L. Friesz, 2015. “Continuity of the Path Delay Operator for LWR-Based Network Loading with Spillback.” Working Paper.
- [15] Heydecker, B.G. and J.D. Addison, 2005. “Analysis of Dynamic Traffic Equilibrium with Departure Time Choice.” *Transportation Science* 39, 39-57.
- [16] Hofbauer, J. and W.H. Sandholm, 2007. “Evolution in Games with Randomly Disturbed Payoffs.” *Journal of Economic Theory* 132, 47-69.
- [17] Hu, D., 2010. “Equilibrium and Dynamics of the Discrete Corridor Problem.” Unpublished manuscript.
- [18] Ito, T., 1979. “A Filippov Solution of a System of Differential Equations with Discontinuous Right-Hand Sides.” *Economics Letters* 4, 349-354.
- [19] Jacobsson, K., 2008. *Dynamic Modeling of Internet Congestion Control*. Doctoral Thesis, KTH School of Electrical Engineering: Stockholm.
- [20] Kelly, F.P, A.K. Maulloo and D.K.H. Tan, 1998. “Rate Control for Communication Networks: Shadow Prices, Proportional Fairness, and Stability.” *Journal of the Operational Research Society* 49, 237-252.
- [21] Khan, M.A., 1989. “On Cournot-Nash Equilibrium Distributions for Games with a Nonmetrizable Action Space and Upper Semicontinuous Payoffs.” *Transactions of the American Mathematical Society* 315, 127-146.
- [22] Konishi, H., 2004. “Uniqueness of User Equilibrium in Transportation Networks with Heterogeneous Commuters.” *Transportation Science* 38, 315-330.

- [23] Marson, A., 2004. “Nonconvex Conservation Laws and Ordinary Differential Equations.” *Journal of the London Mathematical Society* 69, 428-440.
- [24] Merchant, D.K. and G.L. Nemhauser, 1978. “A Model and an Algorithm for the Dynamic Traffic Assignment Problems.” *Transportation Science* 12, 183-199.
- [25] Rosenthal, R.W., 1973. “A Class of Games Possessing Pure-Strategy Nash Equilibria.” *International Journal of Game Theory* 2, 65-67.
- [26] Ross, S.L. and J. Yinger, 2000. “Timing Equilibria in an Urban Model with Congestion.” *Journal of Urban Economics* 47, 390-413.
- [27] Sandholm, W.H., 2001. “Potential Games with Continuous Player Sets.” *Journal of Economic Theory* 97, 81-108.
- [28] Sandholm, W.H., 2007. “Pigouvian Pricing and Stochastic Evolutionary Implementation.” *Journal of Economic Theory* 132, 367–382.
- [29] Schmeidler, D., 1973. “Equilibrium Points of Non-Atomic Games.” *Journal of Statistical Physics* 7, 295-301.
- [30] Smart, D.R., 1974. *Fixed Point Theorems*. Cambridge University Press: Cambridge, UK.
- [31] Strub, I.S. and A.M. Bayen, 2006. “Mixed Initial-Boundary Value Problems for Scalar Conservation Laws: Application to the Modeling of Transportation Networks.” In *Hybrid Systems: Computation and Control*, Lecture Notes in Computer Science Volume 3927, edited by J. Hespanha and A. Tiwari, pp. 552-567. Springer Verlag: Berlin.
- [32] Verhoef, E., 1999. “Time, Speeds, Flows and Densities in Static Models of Road Traffic Congestion and Congestion Pricing.” *Regional Science and Urban Economics* 29, 341-369.
- [33] Vickrey, W., 1963. “Pricing in Urban and Suburban Transport.” *American Economic Review* 53, 452-465.
- [34] Vickrey, W., 1969. “Congestion Theory and Transport Investment.” *American Economic Review* 59, 251-261.

- [35] Zhang, X. and H.M. Zhang, 2010. “Simultaneous Departure Time/Route Choices in Queuing Networks and a Novel Paradox.” *Networks and Spatial Economics* 10, pp. 93-112.
- [36] Zhu, D. and P. Marcotte, 2000. “On the Existence of Solutions to the Dynamic User Equilibrium Problem.” *Transportation Science* 34, 402-414.

5 Appendix: Proofs

5.1 Proof of Theorem 1

Preliminaries: We work to find a unique fixed point in initial conditions g and progress along a link δ . The main issue is consistency of the commuting pattern with boundary values on links, namely the density of departures along a link from a node. These initial conditions are partly exogenous, due to the fixed choice of departure times and routes for Theorem 1 (in contrast with Theorem 2), and partly endogenous, for nodes along a commuter’s route that are not the point of departure. So we employ a fixed point on this data; it will be in a subspace of functions of bounded variation for g .

We have already defined the space where δ lives; see (6). Recalling that each permissible route can only go through a given node once, next we define the space of all possible initial conditions, \mathcal{G} :

Definitions: For $g : [0, \bar{t}] \rightarrow \mathbb{R}_+$, the total variation norm is defined as:

$$TV(g) \equiv \sup \left\{ \sum_{k=1}^K |g(t_k) - g(t_{k-1})| \mid K \geq 1, t_k \in [0, \bar{t}], t_0 < t_1 < \dots < t_K \right\}$$

Next, define

$$\underline{g} \equiv \min \left\{ \Phi_{mn}^{-1} \left(\frac{\mu(\{c \in C \mid R(c) = r, |\tau^d(c) - \tau| < \bar{\tau}\})}{2\bar{\tau}} \right) \mid r \in \mathcal{R}, \tau \in [0, \bar{t}], \pi_1(R(c)) = m, \pi_2(R(c)) = n, \mu(\{c \in C \mid R(c) = r, |\tau^d(c) - \tau| < \bar{\tau}\}) > 0 \right\} > 0$$

Number the equivalence classes of links defined by the relation \succeq , from the bottom class up, using the index $h = 1, 2, \dots, H \leq N^2 - N$. Define $\chi_1 \equiv (T/\bar{\tau} + 1) \cdot \bar{f}$ and define inductively

$$\chi_{h+1} \equiv |\mathcal{R}| \cdot (T/\bar{\tau} + 1) \cdot \bar{f} + (N - 1) \cdot \chi_h \cdot \frac{\bar{f}}{\underline{g}}$$

for $h = 1, 2, \dots, H - 1$. For link mn , define $h(mn)$ as the equivalence class to which it is assigned. Finally, define:

$$\mathcal{G}_m^r \equiv \left\{ \begin{array}{l} \left\{ \Phi_{mn}^{-1} \left(\frac{\mu(\{c \in C \mid R(c)=r, |\tau^d(c)-\tau| < \bar{\tau}\})}{2\bar{\tau}} \right) \right\} \text{ if } \pi_1(r) = m, \pi_2(r) = n; \\ \left\{ g_m^r(\cdot) \text{ measurable on } [0, \bar{t}] \mid 0 \leq g_m^r(\cdot) \leq \bar{f}, g_m^r(\bar{t}) = 0, TV(g_m^r(\cdot)) \leq \chi_{h(mn)}^r \right\} \\ \text{if } \pi_i(r) = m, \pi_{i+1}(r) = n \text{ for some } i > 1 \end{array} \right\}$$

Notice that for Theorem 1, flux Φ_{mn} is one to one.

$$\mathcal{G} \equiv \prod_{n=1}^N \prod_{\{r \in \mathcal{R} \mid \pi_i(r)=n \text{ for some } i\}} \mathcal{G}_n^r$$

We shall be searching for a fixed point in \mathcal{G} . So the next step is to define the map from \mathcal{G} into itself.

We begin by fixing some $g \in \mathcal{G}$, so that for each link mn , $m \neq n$, define

$$g(m, n, \cdot) \equiv \sum_{\{r \in \mathcal{R} \mid \pi_i(r)=m, \pi_{i+1}(r)=n \text{ for some } i \geq 1\}} g_n^r(\cdot)$$

Define

$$\chi_{mn} \equiv (N - 1) \cdot \chi_{h(mn)}$$

Notice that $TV(g(m, n, \cdot)) \leq \chi_{mn}$. After some preparation, we shall define the map $\mathcal{T} : \mathcal{G} \rightarrow \mathcal{G}$. We will call $\mathcal{T}(g) \equiv \hat{g}$. Next we begin preparations for defining this map.

Given the initial condition

$$f(m, n, 0, \Delta) = 0 \quad \forall \Delta \geq 0$$

and the left boundary condition on each link mn , $m \neq n$: $f(m, n, \tau, 0) = g(m, n, \tau)$, Strub and Bayen (2006) yields existence of a unique solution (as we have defined it) called $f(m, n, \tau, \Delta)$. We must be a little careful here, specifically at the right boundary $\lambda(m, n)$. Although they only use the solution on $(0, \bar{t}) \times (0, \lambda(m, n))$, as they remark, it is in fact defined on $[0, \bar{t}] \times [0, \lambda(m, n)]$. All we need is that it is defined on $(0, \bar{t}) \times (0, \lambda(m, n)]$. Second, to make the right boundary condition non-binding, we simply set (in their notation) $\rho_b(t) = 0$. Then the right boundary condition becomes vacuous.³² The initial

³²In fact, this is where we use Assumption 3, implying that there is no backup onto a link of traffic congestion, namely a threshold, at the endpoint of that link. In particular, we ignore behavior outside the link when we solve the differential equation for traffic flow on a link.

(in contrast with the boundary) condition is: at time 0, the density of traffic along the link is 0. Only the left boundary condition will apply in a significant way.

Next we define a unique $\delta_{mn}(\hat{\tau}, \tau) \in \mathcal{D}_{mn}$ associated with f . To accomplish this, we shall apply Bressan (1988) Theorem 1 to the (discontinuous) ordinary differential equation:

$$\frac{\partial \delta_{mn}(\hat{\tau}, \tau)}{\partial \tau} = v(f(m, n, \tau, \delta_{mn}(\hat{\tau}, \tau)), x_{mn}) \quad (10)$$

This will require us to delve a little into the clever proof of existence of a solution f used by Strub and Bayen (2006)³³ in order to integrate it with the structure of Bressan (1988).³⁴ These ideas will also be useful shortly in order to prove that $\hat{g} \in \mathcal{G}$.

There are two (sufficient) conditions for the existence and uniqueness of a solution to the differential equation (10). The first is *locally bounded Γ^M variation*, when specialized to our context is as follows. Let \prec^* be the partial order on \mathbb{R}^2 :

$$(t, \Delta) \prec^* (t', \Delta') \text{ if and only if } |\Delta' - \Delta| \leq M \cdot (t' - t)$$

A vector field $V : \mathbb{R}^2 \rightarrow \mathbb{R}$ is said to have *locally bounded Γ^M variation*³⁵ if, for every $(t_0, x_0) \in \mathbb{R}^2$, there exist $\zeta, \omega > 0$ such that:

$$\sum_{k=1}^K |V(t_k, \Delta_k) - V(t_{k-1}, \Delta_{k-1})| \leq \omega$$

for every finite sequence (t_k, Δ_k) ($k = 1, \dots, K$) with

$$\begin{aligned} (t_0, \Delta_0) &\prec^* (t_1, \Delta_1) \cdots \prec^* (t_K, \Delta_K) \\ t_K &< t_0 + \zeta \end{aligned}$$

The second condition is:

$$|V(t, \Delta)| \leq L < M \text{ for all } (t, \Delta)$$

For our application, take J to be the smallest integer larger than $\max_{m,n} \frac{\bar{f}}{v(\bar{f}, x_{mn})}$. Then $L = \bar{f}$ and $M = J \cdot \max_{m,n} v(\bar{f}, x_{mn})$. We also know from Strub and

³³The keys to this proof are the Godunov construction and the Courant-Friedrichs-Lewy condition.

³⁴Since there are notational conflicts between the two papers as well as with our notation, integration requires some notational changes.

³⁵It is quite amazing to an outsider that this definition seems not to have been related to the standard definition of bounded variation on \mathbb{R}^2 that relies on calculus.

Bayen (2006), p. 560, that for each $\Delta \in [0, \lambda(m, n)]$, $\widehat{f}(m, n, \cdot, \Delta)$ is of bounded variation. But for our purposes, it will be useful to prove the stronger assertion: For each $\Delta \in [0, \lambda(m, n)]$,

$$TV(\widehat{f}(m, n, \cdot, \Delta)) \leq \chi_{mn}.$$

That is next on the agenda.

Strub and Bayen (2006) use an approximation to construct the solution that we call $\widehat{f}(m, n, \cdot, \cdot)$. In their notation, they consider only one link and thus drop m and n . To reduce notation, we also drop these indexes temporarily. The discrete approximation they use is called ρ_i^s , where s denotes a time cell and i denotes a location cell, and where i and s are integers. Specifically, the cells are defined as follows:

$$\begin{aligned} I_i &= \left[\frac{\lambda}{\mathcal{M}} \cdot \left(i - \frac{1}{2}\right), \frac{\lambda}{\mathcal{M}} \cdot \left(i + \frac{1}{2}\right) \right] \\ J_s &= \left[\frac{\lambda}{\mathcal{M}} z \cdot \left(s - \frac{1}{2}\right), \frac{\lambda}{\mathcal{M}} z \cdot \left(s + \frac{1}{2}\right) \right] \end{aligned}$$

where λ is the length of the link, \mathcal{M} denotes the number of location cells ($i = 1, \dots, \mathcal{M}$) and the constant $z > 0$. The cell sizes tend to zero ($\mathcal{M} \rightarrow \infty$) as the approximation converges. It is important to note that, from the uniqueness result, the limit is actually independent of z . The key equation system from Strub and Bayen (2006, p. 559) is as follows:

$$\begin{cases} \rho_{i+\frac{1}{2}}^s \text{ is an element } f \text{ of } I(\rho_i^n, \rho_{i+1}^n) \text{ such that } \text{sign}(\rho_{i+1}^n - \rho_i^n) \cdot \Phi(f) \text{ is minimal} \\ \rho_i^{s+1} = \rho_i^s - z \cdot \left(\Phi(\rho_{i+\frac{1}{2}}^s) - \Phi(\rho_{i-\frac{1}{2}}^s) \right) \end{cases}$$

Our assumptions about the flux function Φ allow us to simplify this. Let ϕ be the Lipschitz constant for Φ . Then

$$\begin{aligned} |\rho_i^{s+1} - \rho_i^s| &= z \cdot \left| \Phi(\rho_{i+\frac{1}{2}}^s) - \Phi(\rho_{i-\frac{1}{2}}^s) \right| \\ &\leq z \cdot \phi \cdot \left| \rho_{i+\frac{1}{2}}^s - \rho_{i-\frac{1}{2}}^s \right| \\ &= z \cdot \phi \cdot |\rho_i^s - \rho_{i-1}^s| \end{aligned}$$

Choosing $z \leq \frac{1}{\phi}$, we obtain:

$$|\rho_i^{s+1} - \rho_i^s| \leq |\rho_i^s - \rho_{i-1}^s| \quad (11)$$

Dropping m and n , if $g(t)$ is the left boundary condition at time t , the left boundary condition for the discrete approximation is given by:

$$\rho_0^s = \frac{\mathcal{M}}{z \cdot \lambda} \int_{J_s} g(t) dt$$

To start, if we fix location $i = 1$ and sum (11) over time s , we can see that the total variation at location cell 1 is bounded by the total variation of ρ_0^s over time s , that is in turn at most the constant χ_{mn} . By induction on i , we can see that this holds for each i . Finally, from Strub and Bayen (2006), these step functions ρ converge strongly (in L^1) to the function \widehat{f} of bounded variation. In our context, for each fixed i , we can apply Helly's theorem to obtain pointwise convergence of a subsequence, implying that the limit \widehat{f} satisfies $TV(\widehat{f}(m, n, \cdot, \Delta)) \leq \chi_{mn}$. To obtain the exit density, we take a sequence $\{\Delta_n\}_{n=1}^\infty$ with $\Delta_n < \lambda(m, n)$ and $\lim_{n \rightarrow \infty} \Delta_n = \lambda(m, n)$. Then again apply Helly's theorem to obtain the pointwise limit of a subsequence, and call this density $\widehat{f}(m, n, \cdot, \lambda(m, n))$.

This exit density (as a function of time) will form the basis for entry density on succeeding links. Notice that $\widehat{f}(m, n, \cdot, \lambda(m, n))$ is of bounded variation. By remark 2.1 of Bressan (2000), we can take it to be right continuous in t .

Next, we examine whether this limit exit density is unique, at least among functions of bounded variation that satisfy $TV(f(m, n, \cdot, \Delta)) \leq \chi_{mn}$. Suppose that there are two different exit limits of bounded variation; call them $\widetilde{f}(m, n, \cdot, \lambda(m, n))$ and $\widehat{f}(m, n, \cdot, \lambda(m, n))$. Now we already know from Strub and Bayen (2006) that $\widetilde{f}(m, n, \cdot, \cdot) = \widehat{f}(m, n, \cdot, \cdot)$ a.s. (t, Δ) . The next argument parallels Strub and Bayen (pp. 558-559) where they argue that their solution is unique. We also know for $\varphi \in C_c^1(0, \bar{t})$ and $\psi \in C_c^1(0, \lambda(m, n))$ (where $\varphi, \psi \geq 0$),

$$\begin{aligned} & \int_0^{\lambda(m, n)} \int_0^{\bar{t}} \left| \widetilde{f}(m, n, t, \Delta) - \widehat{f}(m, n, t, \Delta) \right| \psi(\Delta) \varphi'(t) \\ & + \text{sign} \left(\widetilde{f}(m, n, t, \Delta) - \widehat{f}(m, n, t, \Delta) \right) \cdot \left(\Phi \left(\widetilde{f}(m, n, t, \Delta) \right) - \Phi \left(\widehat{f}(m, n, t, \Delta) \right) \right) \cdot \psi'(\Delta) \varphi(t) d\Delta dt \\ & \geq 0 \end{aligned}$$

For φ approximating the indicator function of $[0, \bar{t}]$, we have:

$$\begin{aligned} & \limsup_{t \rightarrow 0} \int_0^{\lambda(m, n)} \left| \widetilde{f}(m, n, t, \Delta) - \widehat{f}(m, n, t, \Delta) \right| \psi(\Delta) dt \\ & - \liminf_{t \rightarrow \bar{t}} \int_0^{\lambda(m, n)} \left| \widetilde{f}(m, n, t, \Delta) - \widehat{f}(m, n, t, \Delta) \right| \psi(\Delta) dt \\ & \geq - \int_0^{\lambda(m, n)} \int_0^{\bar{t}} \text{sign} \left(\widetilde{f}(m, n, t, \Delta) - \widehat{f}(m, n, t, \Delta) \right) \cdot \\ & \left(\Phi \left(\widetilde{f}(m, n, t, \Delta) \right) - \Phi \left(\widehat{f}(m, n, t, \Delta) \right) \right) \cdot \psi'(\Delta) d\Delta dt \end{aligned}$$

Taking ψ to approximate the indicator function of $[0, \lambda(m, n)]$,

$$\begin{aligned}
&\geq \limsup_{\Delta \rightarrow \lambda(m, n)} \int_0^{\bar{t}} \text{sign} \left(\tilde{f}(m, n, t, \Delta) - \hat{f}(m, n, t, \Delta) \right) \cdot \\
&\quad \left(\Phi \left(\tilde{f}(m, n, t, \Delta) \right) - \Phi \left(\hat{f}(m, n, t, \Delta) \right) \right) dt \\
&\quad - \liminf_{\Delta \rightarrow 0} \int_0^{\bar{t}} \text{sign} \left(\tilde{f}(m, n, t, \Delta) - \hat{f}(m, n, t, \Delta) \right) \cdot \\
&\quad \left(\Phi \left(\tilde{f}(m, n, t, \Delta) \right) - \Phi \left(\hat{f}(m, n, t, \Delta) \right) \right) dt
\end{aligned}$$

In sum, we have:

$$\begin{aligned}
&\limsup_{t \rightarrow 0} \int_0^{\lambda(m, n)} \left| \tilde{f}(m, n, t, \Delta) - \hat{f}(m, n, t, \Delta) \right| \psi(\Delta) dt - \\
&\liminf_{t \rightarrow \bar{t}} \int_0^{\lambda(m, n)} \left| \tilde{f}(m, n, t, \Delta) - \hat{f}(m, n, t, \Delta) \right| \psi(\Delta) dt \\
&\geq \limsup_{\Delta \rightarrow \lambda(m, n)} \int_0^{\bar{t}} \text{sign} \left(\tilde{f}(m, n, t, \Delta) - \hat{f}(m, n, t, \Delta) \right) \cdot \\
&\quad \left(\Phi \left(\tilde{f}(m, n, t, \Delta) \right) - \Phi \left(\hat{f}(m, n, t, \Delta) \right) \right) dt \\
&\quad - \liminf_{\Delta \rightarrow 0} \int_0^{\bar{t}} \text{sign} \left(\tilde{f}(m, n, t, \Delta) - \hat{f}(m, n, t, \Delta) \right) \cdot \\
&\quad \left(\Phi \left(\tilde{f}(m, n, t, \Delta) \right) - \Phi \left(\hat{f}(m, n, t, \Delta) \right) \right) dt
\end{aligned}$$

Since the left hand side (the first two terms) are zero, we obtain:

$$\begin{aligned}
&\liminf_{\Delta \rightarrow 0} \int_0^{\bar{t}} \text{sign} \left(\tilde{f}(m, n, t, \Delta) - \hat{f}(m, n, t, \Delta) \right) \cdot \\
&\quad \left(\Phi \left(\tilde{f}(m, n, t, \Delta) \right) - \Phi \left(\hat{f}(m, n, t, \Delta) \right) \right) dt \\
&\geq \limsup_{\Delta \rightarrow \lambda(m, n)} \int_0^{\bar{t}} \text{sign} \left(\tilde{f}(m, n, t, \Delta) - \hat{f}(m, n, t, \Delta) \right) \cdot \\
&\quad \left(\Phi \left(\tilde{f}(m, n, t, \Delta) \right) - \Phi \left(\hat{f}(m, n, t, \Delta) \right) \right) dt
\end{aligned}$$

As in Strub and Bayen (2006, p. 558), the left hand side is 0. The right hand side is non-negative (recall that flux Φ is non-decreasing in density). Hence,

$$\begin{aligned}
&\limsup_{\Delta \rightarrow \lambda(m, n)} \int_0^{\bar{t}} \text{sign} \left(\tilde{f}(m, n, t, \Delta) - \hat{f}(m, n, t, \Delta) \right) \cdot \left(\Phi \left(\tilde{f}(m, n, t, \Delta) \right) - \Phi \left(\hat{f}(m, n, t, \Delta) \right) \right) dt \\
&\quad = 0
\end{aligned}$$

Now since Φ is strictly increasing in f , then we know that

$$\lim_{\Delta \rightarrow \lambda(m,n)} \sup \left\| f(m, n, \cdot, \Delta) - \widehat{f}(m, n, \cdot, \Delta) \right\|_{L^1} = 0,$$

implying that $\widetilde{f}(m, n, \cdot, \lambda(m, n)) = \widehat{f}(m, n, \cdot, \lambda(m, n))$ a.s. (t) . Both $\widetilde{f}(m, n, \cdot, \lambda(m, n))$ and $\widehat{f}(m, n, \cdot, \lambda(m, n))$ are of bounded variation, so by Lemma 2.1 and Remark 2.1 of Bressan (2000), by taking right continuous versions, they are in fact equal.

Although we used the argument just above to obtain a well-defined exit density, if we replace $\lambda(m, n)$ with an arbitrary distance $\overline{\Delta}$, $0 < \overline{\Delta} < \lambda(m, n)$, the same argument applies and we have that for any sequence $\{\overline{\Delta}_k\}_{k=1}^{\infty}$ with $\lim_{k \rightarrow \infty} \overline{\Delta}_k = \overline{\Delta}$, $\lim_{k \rightarrow \infty} \widehat{f}(m, n, \cdot, \overline{\Delta}_k) = \widehat{f}(m, n, \cdot, \overline{\Delta})$ a.s. (t) , where

$$TV(\widehat{f}(m, n, \cdot, \overline{\Delta})) \leq \chi_{mn}.$$

Taking the right continuous version, it follows that $\lim_{k \rightarrow \infty} \widehat{f}(m, n, \cdot, \overline{\Delta}_k) = \widehat{f}(m, n, \cdot, \overline{\Delta})$; for if not, then by Helly's theorem there are at least two (point-wise) limits and $\widehat{f}(m, n, \cdot, \overline{\Delta})$ is not well-defined, a contradiction.

Now fix (t_0, Δ_0) and take $p = 1, 2, \dots$, and a finite sequence (t_k, Δ_k) ($k = 1, \dots, K$) with

$$\begin{aligned} (t_0, \Delta_0) &\prec^* (t_1, \Delta_1) \cdots \prec^* (t_K, \Delta_K) \\ t_K &< t_0 + \frac{1}{p} \end{aligned}$$

In order to apply Bressan (1988), we must show that for any such sequence, there exist ω and p (that can depend on (t_0, Δ_0)) such that:

$$\sum_{k=1}^K |v(f(m, n, t_k, \Delta_k), x_{mn}) - v(f(m, n, t_{k-1}, \Delta_{k-1}), x_{mn})| \leq \omega \quad (12)$$

To ease notation and break the proof down a little, we first prove:

$$\sum_{k=1}^K |f(m, n, t_k, \Delta_k) - f(m, n, t_{k-1}, \Delta_{k-1})| \leq \omega' \quad (13)$$

Fix $\varepsilon > 0$ and let p be so large that for all k ,

$$TV(f(m, n, \cdot, \Delta_k) - f(m, n, \cdot, \Delta_0)) < \varepsilon$$

With this in hand, since $\{f(m, n, \cdot, \Delta_k)\}_{k=1}^\infty$ is a Cauchy sequence in BV :

$$\begin{aligned}
& \sum_{k=1}^K |f(m, n, t_k, \Delta_k) - f(m, n, t_{k-1}, \Delta_{k-1})| \\
& \leq \sum_{k=1}^K |f(m, n, t_k, \Delta_k) - f(m, n, t_{k-1}, \Delta_k)| + |f(m, n, t_{k-1}, \Delta_k) - f(m, n, t_{k-1}, \Delta_{k-1})| \\
& = \sum_{k=1}^K |f(m, n, t_k, \Delta_k) - f(m, n, t_{k-1}, \Delta_k)| + \sum_{k=1}^K |f(m, n, t_{k-1}, \Delta_k) - f(m, n, t_{k-1}, \Delta_{k-1})| \\
& \leq \sum_{k=1}^K |f(m, n, t_k, \Delta_k) - f(m, n, t_{k-1}, \Delta_k)| + TV(f(m, n, \cdot, \Delta_k) - f(m, n, \cdot, \Delta_{k-1})) \\
& \leq TV(f(m, n, \cdot, \Delta_k)) + 2\varepsilon \\
& \leq \chi_{mn} + 2\varepsilon
\end{aligned}$$

So setting $\omega' > \chi_{mn}$, we have proved the intermediate step (13). To prove (12) from this, we use the fact that v is Lipschitz with constant ϖ , so we set $\omega = \varpi \cdot \omega'$. Thus, we have fulfilled the assumptions of Bressan (1988), Theorem 1, so there is a unique (forward) solution $\delta_{mn}(\hat{\tau}, \tau)$ to (10) for any given f .

Next, we apply the arguments elaborated above to define, and discover properties of, the map $\mathcal{T} : \mathcal{G} \rightarrow \mathcal{G}$. Let $f(m, n, \cdot, \cdot)$ be the unique solution to the conservation law on link mn with initial condition 0 and boundary condition $g(m, n, \cdot)$.

Let $\delta_{mn}(\hat{\tau}, \tau)$ be the corresponding (unique) solution to the differential equation (10). Define:

$$\hat{\tau}_{mn}(\tau) = \delta_{mn}^{-1}(\cdot, \tau)(\lambda(m, n)) \quad (14)$$

Notice that since speed $v > 0$, $\delta_{mn}(\hat{\tau}, \tau)$ is strictly decreasing in $\hat{\tau}$, so $\hat{\tau}_{mn}(\tau)$ is well-defined.

With this preparation, we can define the image $\mathcal{T}(g) = \hat{g}$, that will depend on both g , through the solution on a link f as defined above, and δ , through its inverse image $\hat{\tau}$.

$$\hat{g}_n(\tau) \equiv \begin{cases} \frac{g_m^r(\hat{\tau}_{mn}(\tau))}{\sum_{\{r' \in \mathcal{R} \mid \text{for some } i \ \pi_i(r')=m, \ \pi_{i+1}(r')=n\}} g_m^{r'}(\hat{\tau}_{mn}(\tau))} \cdot \hat{f}(m, n, \tau, \lambda(m, n)) & \text{if } \exists i \geq 3 \text{ with } \pi_{i-1}(r) = m, \ \pi_i(r) = n \\ \Phi_{mn}^{-1} \left(\frac{\mu(\{c \in C \mid R(c)=r, \ |\tau^d(c)-\tau| < \bar{\tau}\})}{2\bar{\tau}} \right) & \text{if } \pi_1(r) = m, \ \pi_2(r) = n \end{cases} \quad (15)$$

The argument that $\hat{g} \in \mathcal{G}$ is as follows.

First, for the case $\pi_1(r) = m, \pi_2(r) = n$, by definition

$$TV(\widehat{g}_n^r(\tau)) = TV\left(\Phi_{mn}^{-1}\left(\frac{\mu(\{c \in C \mid \pi_1(R(c)) = n, |\tau^d(c) - \tau| < \bar{\tau}\})}{2\bar{\tau}}\right)\right) \leq (T/\bar{\tau}+1) \cdot \bar{f}$$

In all other cases,

$$\begin{aligned} TV(\widehat{g}_n^r(\cdot)) &= TV\left(\frac{g_m^r(\widehat{\tau}_{mn}(\tau))}{\sum_{\{r' \in \mathcal{R} \mid \text{for some } i \pi_i(r')=m, \pi_{i+1}(r')=n\}} g_m^{r'}(\widehat{\tau}_{mn}(\tau))} \cdot \widehat{f}(m, n, \tau, \lambda(m, n))\right) \\ &= \sup_{K \geq 1, t_k \in [0, \bar{t}], t_0 < t_1 < \dots < t_K} \left\{ \sum_{k=1}^K \left| \frac{g_m^r(\widehat{\tau}_{mn}(t_k))}{\sum_{\{r' \in \mathcal{R} \mid \text{for some } i \pi_i(r')=m, \pi_{i+1}(r')=n\}} g_m^{r'}(\widehat{\tau}_{mn}(t_k))} \cdot \widehat{f}(m, n, t_k, \lambda(m, n)) \right. \right. \\ &\quad \left. \left. - \frac{g_m^r(\widehat{\tau}_{mn}(t_{k-1}))}{\sum_{\{r' \in \mathcal{R} \mid \text{for some } i \pi_i(r')=m, \pi_{i+1}(r')=n\}} g_m^{r'}(\widehat{\tau}_{mn}(t_{k-1}))} \cdot \widehat{f}(m, n, t_{k-1}, \lambda(m, n)) \right| \right\} \\ &\leq \sup_{K \geq 1, t_k \in [0, \bar{t}], t_0 < t_1 < \dots < t_K} \left\{ \sum_{k=1}^K \frac{g_m^r(\widehat{\tau}_{mn}(t_k))}{\sum_{\{r' \in \mathcal{R} \mid \text{for some } i \pi_i(r')=m, \pi_{i+1}(r')=n\}} g_m^{r'}(\widehat{\tau}_{mn}(t_k))} \right. \\ &\quad \left. \cdot \left| \widehat{f}(m, n, t_k, \lambda(m, n)) - \widehat{f}(m, n, t_{k-1}, \lambda(m, n)) \right| \right\} \\ &+ \sup_{K \geq 1, t_k \in [0, \bar{t}], t_0 < t_1 < \dots < t_K} \left\{ \sum_{k=1}^K \left| \frac{g_m^r(\widehat{\tau}_{mn}(t_k))}{\sum_{\{r' \in \mathcal{R} \mid \text{for some } i \pi_i(r')=m, \pi_{i+1}(r')=n\}} g_m^{r'}(\widehat{\tau}_{mn}(t_k))} \right. \right. \\ &\quad \left. \left. - \frac{g_m^r(\widehat{\tau}_{mn}(t_{k-1}))}{\sum_{\{r' \in \mathcal{R} \mid \text{for some } i \pi_i(r')=m, \pi_{i+1}(r')=n\}} g_m^{r'}(\widehat{\tau}_{mn}(t_{k-1}))} \right| \cdot \widehat{f}(m, n, t_{k-1}, \lambda(m, n)) \right\} \\ &\leq TV(\widehat{f}(m, n, \cdot, \lambda(m, n))) \\ &+ \sup_{K \geq 1, t_k \in [0, \bar{t}], t_0 < t_1 < \dots < t_K} \left\{ \sum_{k=1}^K \left| \frac{g_m^r(\widehat{\tau}_{mn}(t_k))}{\sum_{\{r' \in \mathcal{R} \mid \text{for some } i \pi_i(r')=m, \pi_{i+1}(r')=n\}} g_m^{r'}(\widehat{\tau}_{mn}(t_k))} \right. \right. \\ &\quad \left. \left. - \frac{g_m^r(\widehat{\tau}_{mn}(t_{k-1}))}{\sum_{\{r' \in \mathcal{R} \mid \text{for some } i \pi_i(r')=m, \pi_{i+1}(r')=n\}} g_m^{r'}(\widehat{\tau}_{mn}(t_{k-1}))} \right| \right\} \cdot \bar{f} \quad (16) \end{aligned}$$

To simplify this expression further, we focus on the second term. For notational brevity, define:

$$\Xi \equiv \frac{1}{\left[\sum_{\{r' \in \mathcal{R} \mid \exists i \text{ s.t. } \pi_i(r')=m, \pi_{i+1}(r')=n\}} g_m^{r'}(\widehat{\tau}_{mn}(t_k)) \right] \cdot \left[\sum_{\{r' \in \mathcal{R} \mid \exists i \text{ s.t. } \pi_i(r')=m, \pi_{i+1}(r')=n\}} g_m^{r'}(\widehat{\tau}_{mn}(t_{k-1})) \right]}$$

$$\begin{aligned}
& \left| \frac{g_m^r(\widehat{\tau}_{mn}(t_k))}{\sum_{\{r' \in \mathcal{R} | \text{for some } i \pi_i(r')=m, \pi_{i+1}(r')=n\}} g_m^{r'}(\widehat{\tau}_{mn}(t_k))} \right. \\
& \quad \left. - \frac{g_m^r(\widehat{\tau}_{mn}(t_{k-1}))}{\sum_{\{r' \in \mathcal{R} | \text{for some } i \pi_i(r')=m, \pi_{i+1}(r')=n\}} g_m^{r'}(\widehat{\tau}_{mn}(t_{k-1}))} \right| \\
&= \Xi \cdot \left| \left[\sum_{\{r' \in \mathcal{R} | \text{for some } i \pi_i(r')=m, \pi_{i+1}(r')=n\}} g_m^{r'}(\widehat{\tau}_{mn}(t_{k-1})) \right] \cdot g_m^r(\widehat{\tau}_{mn}(t_k)) \right. \\
& \quad \left. - g_m^r(\widehat{\tau}_{mn}(t_{k-1})) \cdot \left[\sum_{\{r' \in \mathcal{R} | \text{for some } i \pi_i(r')=m, \pi_{i+1}(r')=n\}} g_m^{r'}(\widehat{\tau}_{mn}(t_k)) \right] \right| \\
&\leq \Xi \cdot \left(\left[\sum_{\{r' \in \mathcal{R} | \text{for some } i \pi_i(r')=m, \pi_{i+1}(r')=n\}} g_m^{r'}(\widehat{\tau}_{mn}(t_{k-1})) \right] \cdot |g_m^r(\widehat{\tau}_{mn}(t_k)) - g_m^r(\widehat{\tau}_{mn}(t_{k-1}))| \right. \\
& \quad \left. + g_m^r(\widehat{\tau}_{mn}(t_{k-1})) \cdot \left| \left[\sum_{\{r' \in \mathcal{R} | \text{for some } i \pi_i(r')=m, \pi_{i+1}(r')=n\}} g_m^{r'}(\widehat{\tau}_{mn}(t_{k-1})) \right] \right. \right. \\
& \quad \left. \left. - \left[\sum_{\{r' \in \mathcal{R} | \text{for some } i \pi_i(r')=m, \pi_{i+1}(r')=n\}} g_m^{r'}(\widehat{\tau}_{mn}(t_k)) \right] \right| \right) \\
&\leq \frac{|g_m^r(\widehat{\tau}_{mn}(t_k)) - g_m^r(\widehat{\tau}_{mn}(t_{k-1}))|}{\left[\sum_{\{r' \in \mathcal{R} | \text{for some } i \pi_i(r')=m, \pi_{i+1}(r')=n\}} g_m^{r'}(\widehat{\tau}_{mn}(t_k)) \right]} \\
& \quad + \frac{1}{\left[\sum_{\{r' \in \mathcal{R} | \text{for some } i \pi_i(r')=m, \pi_{i+1}(r')=n\}} g_m^{r'}(\widehat{\tau}_{mn}(t_k)) \right]} \cdot \\
& \quad \left| \left[\sum_{\{r' \in \mathcal{R} | \text{for some } i \pi_i(r')=m, \pi_{i+1}(r')=n\}} g_m^{r'}(\widehat{\tau}_{mn}(t_{k-1})) \right] \right. \\
& \quad \left. - \left[\sum_{\{r' \in \mathcal{R} | \text{for some } i \pi_i(r')=m, \pi_{i+1}(r')=n\}} g_m^{r'}(\widehat{\tau}_{mn}(t_k)) \right] \right| \\
& \quad \cdot \left(\frac{1}{\left[\sum_{\{r' \in \mathcal{R} | \text{for some } i \pi_i(r')=m, \pi_{i+1}(r')=n\}} g_m^{r'}(\widehat{\tau}_{mn}(t_k)) \right]} \cdot (|g_m^r(\widehat{\tau}_{mn}(t_k)) - g_m^r(\widehat{\tau}_{mn}(t_{k-1}))| + \right. \\
& \quad \left. \left| \left[\sum_{\{r' \in \mathcal{R} | \text{for some } i \pi_i(r')=m, \pi_{i+1}(r')=n\}} g_m^{r'}(\widehat{\tau}_{mn}(t_{k-1})) \right] \right. \right. \\
& \quad \left. \left. - \left[\sum_{\{r' \in \mathcal{R} | \text{for some } i \pi_i(r')=m, \pi_{i+1}(r')=n\}} g_m^{r'}(\widehat{\tau}_{mn}(t_k)) \right] \right| \right)
\end{aligned}$$

Similarly,

$$\begin{aligned}
& \left| \frac{g_m^r(\widehat{\tau}_{mn}(t_k))}{\sum_{\{r' \in \mathcal{R} \mid \text{for some } i \ \pi_i(r')=m, \ \pi_{i+1}(r')=n\}} g_m^{r'}(\widehat{\tau}_{mn}(t_k))} \right| \\
& - \left| \frac{g_m^r(\widehat{\tau}_{mn}(t_{k-1}))}{\sum_{\{r' \in \mathcal{R} \mid \text{for some } i \ \pi_i(r')=m, \ \pi_{i+1}(r')=n\}} g_m^{r'}(\widehat{\tau}_{mn}(t_{k-1}))} \right| \\
& \leq \frac{1}{\left[\sum_{\{r' \in \mathcal{R} \mid \text{for some } i \ \pi_i(r')=m, \ \pi_{i+1}(r')=n\}} g_m^{r'}(\widehat{\tau}_{mn}(t_{k-1})) \right]} \cdot \left(|g_m^r(\widehat{\tau}_{mn}(t_k)) - g_m^r(\widehat{\tau}_{mn}(t_{k-1}))| \right. \\
& + \left| \left[\sum_{\{r' \in \mathcal{R} \mid \text{for some } i \ \pi_i(r')=m, \ \pi_{i+1}(r')=n\}} g_m^{r'}(\widehat{\tau}_{mn}(t_{k-1})) \right] \right. \\
& - \left. \left. \left[\sum_{\{r' \in \mathcal{R} \mid \text{for some } i \ \pi_i(r')=m, \ \pi_{i+1}(r')=n\}} g_m^{r'}(\widehat{\tau}_{mn}(t_k)) \right] \right| \right)
\end{aligned}$$

Hence,

$$\begin{aligned}
& \left| \frac{g_m^r(\widehat{\tau}_{mn}(t_k))}{\sum_{\{r' \in \mathcal{R} \mid \text{for some } i \ \pi_i(r')=m, \ \pi_{i+1}(r')=n\}} g_m^{r'}(\widehat{\tau}_{mn}(t_k))} \right| \\
& - \left| \frac{g_m^r(\widehat{\tau}_{mn}(t_{k-1}))}{\sum_{\{r' \in \mathcal{R} \mid \text{for some } i \ \pi_i(r')=m, \ \pi_{i+1}(r')=n\}} g_m^{r'}(\widehat{\tau}_{mn}(t_{k-1}))} \right| \\
& \leq \left(\max \left\{ \left[\sum_{\{r' \in \mathcal{R} \mid \text{for some } i \ \pi_i(r')=m, \ \pi_{i+1}(r')=n\}} g_m^{r'}(\widehat{\tau}_{mn}(t_k)) \right], \right. \right. \\
& \left. \left. \left[\sum_{\{r' \in \mathcal{R} \mid \text{for some } i \ \pi_i(r')=m, \ \pi_{i+1}(r')=n\}} g_m^{r'}(\widehat{\tau}_{mn}(t_{k-1})) \right] \right\} \right)^{-1} \\
& \cdot \left(|g_m^r(\widehat{\tau}_{mn}(t_{k-1})) - g_m^r(\widehat{\tau}_{mn}(t_k))| + \left| \left[\sum_{\{r' \in \mathcal{R} \mid \text{for some } i \ \pi_i(r')=m, \ \pi_{i+1}(r')=n\}} g_m^{r'}(\widehat{\tau}_{mn}(t_k)) \right] \right. \right. \\
& - \left. \left. \left[\sum_{\{r' \in \mathcal{R} \mid \text{for some } i \ \pi_i(r')=m, \ \pi_{i+1}(r')=n\}} g_m^{r'}(\widehat{\tau}_{mn}(t_{k-1})) \right] \right| \right)
\end{aligned}$$

The key point from the last two expressions is that as long as

$$\sum_{\{r' \in \mathcal{R} \mid \text{for some } i \ \pi_i(r')=m, \ \pi_{i+1}(r')=n\}} g_m^{r'}(\widehat{\tau}_{mn}(t_k)) > 0$$

or

$$\sum_{\{r' \in \mathcal{R} | \text{for some } i \pi_i(r')=m, \pi_{i+1}(r')=n\}} g_m^{r'}(\widehat{\tau}_{mn}(t_{k-1})) > 0,$$

then

$$\max \left\{ \left[\sum_{\{r' \in \mathcal{R} | \text{for some } i \pi_i(r')=m, \pi_{i+1}(r')=n\}} g_m^{r'}(\widehat{\tau}_{mn}(t_k)) \right], \right. \\ \left. \left[\sum_{\{r' \in \mathcal{R} | \text{for some } i \pi_i(r')=m, \pi_{i+1}(r')=n\}} g_m^{r'}(\widehat{\tau}_{mn}(t_{k-1})) \right] \right\} \geq \underline{g}$$

If both are 0, then we can ignore this term in the calculations of $TV(\widehat{g}_n^r(\widehat{\tau}_{mn}(\cdot)))$ and $TV\left(\sum_{\{r' \in \mathcal{R} | \text{for some } i \pi_i(r')=m, \pi_{i+1}(r')=n\}} g_m^{r'}(\widehat{\tau}_{mn}(\cdot))\right)$, since this term is 0.

Therefore, from (16),

$$\begin{aligned} & TV(\widehat{g}_n^r(\cdot)) \leq TV(f(m, n, \cdot, \lambda(m, n))) \\ & + \sup_{K \geq 1, t_k \in [0, \bar{t}], t_0 < t_1 < \dots < t_K} \left\{ \sum_{k=1}^K \left| \frac{g_m^r(\widehat{\tau}_{mn}(t_k))}{\sum_{\{r' \in \mathcal{R} | \text{for some } i \pi_i(r')=m, \pi_{i+1}(r')=n\}} g_m^{r'}(\widehat{\tau}_{mn}(t_k))} \right. \right. \\ & \quad \left. \left. - \frac{g_m^r(\widehat{\tau}_{mn}(t_{k-1}))}{\sum_{\{r' \in \mathcal{R} | \text{for some } i \pi_i(r')=m, \pi_{i+1}(r')=n\}} g_m^{r'}(\widehat{\tau}_{mn}(t_{k-1}))} \right| \right\} \cdot \bar{f} \\ & \leq |\mathcal{R}| \cdot (T/\bar{\tau} + 1) \cdot \bar{f} \\ & + \sup \left\{ \left(\max \left\{ \left[\sum_{\{r' \in \mathcal{R} | \text{for some } i \pi_i(r')=m, \pi_{i+1}(r')=n\}} g_m^{r'}(\widehat{\tau}_{mn}(t_k)) \right], \right. \right. \right. \\ & \quad \left. \left. \left[\sum_{\{r' \in \mathcal{R} | \text{for some } i \pi_i(r')=m, \pi_{i+1}(r')=n\}} g_m^{r'}(\widehat{\tau}_{mn}(t_{k-1})) \right] \right\} \right)^{-1} \\ & \quad \left(\sum_{k=1}^K |g_m^r(\widehat{\tau}_{mn}(t_{k-1})) - g_m^r(\widehat{\tau}_{mn}(t_k))| \right. \\ & \quad \left. + \left| \left[\sum_{\{r' \in \mathcal{R} | \text{for some } i \pi_i(r')=m, \pi_{i+1}(r')=n\}} g_m^{r'}(\widehat{\tau}_{mn}(t_k)) \right] \right. \right. \\ & \quad \left. \left. - \left[\sum_{\{r' \in \mathcal{R} | \text{for some } i \pi_i(r')=m, \pi_{i+1}(r')=n\}} g_m^{r'}(\widehat{\tau}_{mn}(t_{k-1})) \right] \right| \right) \\ & \quad \left. \mid K \geq 1, t_k \in [0, \bar{t}], t_0 < t_1 < \dots < t_K \right\} \cdot \bar{f} \end{aligned}$$

$$\begin{aligned}
& \leq |\mathcal{R}| \cdot (T/\bar{\tau} + 1) \cdot \bar{f} \\
& + \left[TV(g_m^r(\widehat{\tau}_{mn}(\cdot))) + TV \left(\sum_{\{r' \in \mathcal{R} | \text{for some } i \pi_i(r')=m, \pi_{i+1}(r')=n\}} g_m^{r'}(\widehat{\tau}_{mn}(\cdot)) \right) \right] \cdot \frac{\bar{f}}{\underline{g}} \\
& \leq |\mathcal{R}| \cdot (T/\bar{\tau} + 1) \cdot \bar{f} + \left[TV(g_m^r(\cdot)) + TV \left(\sum_{\{r' \in \mathcal{R} | \text{for some } i \pi_i(r')=m, \pi_{i+1}(r')=n\}} g_m^{r'}(\cdot) \right) \right] \cdot \frac{\bar{f}}{\underline{g}}
\end{aligned}$$

To bound this expression, note that if $\pi_1(r) = m$, $TV(g_m^r(\cdot)) \leq (T/\bar{\tau} + 1) \cdot \bar{f} = \chi_1$. More generally, if there is some i with $\pi_i(r) = m$, $\pi_{i+1}(r) = n$, $TV(g_m^r(\cdot)) \leq \chi_{h(mn)}$ and $TV \left(\sum_{\{r' \in \mathcal{R} | \text{for some } i \pi_i(r')=m, \pi_{i+1}(r')=n\}} g_m^{r'}(\cdot) \right) \leq \chi_{mn}$. Hence for any $r \in \mathcal{R}$ with $\pi_{i+2}(r) = n'$, $TV(\widehat{g}_n^r(\cdot)) \leq \chi_{h(nn')}$ and $\widehat{g} \in \mathcal{G}$.

The set \mathcal{G} is obviously convex as a product of convex sets. Imposing the L^1 norm topology on each component \mathcal{G}_n^r , Helly's theorem implies that \mathcal{G}_n^r is compact and hence \mathcal{G} is compact as a product of compact sets. What remains is to show that \mathcal{T} is continuous. Here we use intensively its definition (15).

Let $\{g(\cdot)_q\}_{q=1}^\infty \subseteq \mathcal{G}$ where $\lim_{q \rightarrow \infty} g(\cdot)_q = g(\cdot)$, and thus $\lim_{q \rightarrow \infty} g_n^r(\cdot)_q = g_n^r(\cdot)$ for all $r \in \mathcal{R}$ and n such that $\pi_i(r) = n$ for some i . We must show that $\lim_{q \rightarrow \infty} \mathcal{T}(g(\cdot)_q) = \mathcal{T}(g(\cdot))$. To prove this, we must examine each node in each admissible route independently. So let us focus on node n (subscript) in route r (superscript) for the calculations.

Let $\widehat{f}(m, n, \cdot, \cdot)_q$ be the (unique) solution to the boundary value problem with initial conditions given by

$$g(m, n, \cdot)_q \equiv \sum_{\{r \in \mathcal{R} | \pi_i(r)=m, \pi_{i+1}(r)=n \text{ for some } i \geq 1\}} g_n^r(\cdot)_q$$

Let $\widehat{\tau}_{mn}(\tau)_q$ be the corresponding solution to (14). Next we show that in L^1 , $\widehat{f}(m, n, \cdot, \cdot) = \lim_{q \rightarrow \infty} \widehat{f}(m, n, \cdot, \cdot)_q$ exists and is a solution at initial conditions $g(\cdot)$. The proof traces back through Strub and Bayen's (2006) proof that a solution exists, detailed above, and uses an interchange of limits. The boundary condition at each link mn for the Godunov approximation is given by:

$$\begin{aligned}
\rho_{0,q}^s &= \frac{\mathcal{M}}{\lambda(m, n)} \int_{J_s} g(m, n, \cdot)_q dt \\
\rho_0^s &= \frac{\mathcal{M}}{\lambda(m, n)} \int_{J_s} g(m, n, \cdot) dt
\end{aligned}$$

Evidently, $\rho_{0,q}^s \rightarrow \rho_0^s$. All of the pieces of the proof in Strub and Bayen (2006) rely on $\rho_{0,q}^s$ as well as equalities or weak inequalities. So if they hold for every

element of the sequence, they also hold for the limit. Thus, $\widehat{f}(m, n, \cdot, \cdot) = \lim_{q \rightarrow \infty} \widehat{f}(m, n, \cdot, \cdot)_q$ exists and is the (unique) solution at initial conditions $g(\cdot)$.

$$\begin{aligned}
& \int_0^{\bar{t}} \left| \frac{g_m^r(\widehat{\tau}_{mn}(\tau))}{\sum_{\{r' \in \mathcal{R} \mid \text{for some } i \pi_i(r')=m, \pi_{i+1}(r')=n\}} g_m^{r'}(\widehat{\tau}_{mn}(\tau))} \cdot \widehat{f}(m, n, \tau, \lambda(m, n)) \right. \\
& \quad \left. - \frac{g_m^r(\widehat{\tau}_{mn}(\tau)_q)_q}{\sum_{\{r' \in \mathcal{R} \mid \text{for some } i \pi_i(r')=m, \pi_{i+1}(r')=n\}} g_m^{r'}(\widehat{\tau}_{mn}(\tau)_q)_q} \cdot \widehat{f}(m, n, \tau, \lambda(m, n))_q \right| d\tau \\
& \leq \int_0^{\bar{t}} \left| \frac{g_m^r(\widehat{\tau}_{mn}(\tau))}{\sum_{\{r' \in \mathcal{R} \mid \text{for some } i \pi_i(r')=m, \pi_{i+1}(r')=n\}} g_m^{r'}(\widehat{\tau}_{mn}(\tau))} \cdot f(m, n, \tau, \lambda(m, n)) \right. \\
& \quad \left. - \frac{g_m^r(\widehat{\tau}_{mn}(\tau)_q)_q}{\sum_{\{r' \in \mathcal{R} \mid \text{for some } i \pi_i(r')=m, \pi_{i+1}(r')=n\}} g_m^{r'}(\widehat{\tau}_{mn}(\tau)_q)_q} \cdot \widehat{f}(m, n, \tau, \lambda(m, n)) \right| \\
& \quad + \left| \frac{g_m^r(\widehat{\tau}_{mn}(\tau)_q)_q}{\sum_{\{r' \in \mathcal{R} \mid \text{for some } i \pi_i(r')=m, \pi_{i+1}(r')=n\}} g_m^{r'}(\widehat{\tau}_{mn}(\tau)_q)_q} \cdot \widehat{f}(m, n, \tau, \lambda(m, n)) \right. \\
& \quad \left. - \frac{g_m^r(\widehat{\tau}_{mn}(\tau))_q}{\sum_{\{r' \in \mathcal{R} \mid \text{for some } i \pi_i(r')=m, \pi_{i+1}(r')=n\}} g_m^{r'}(\widehat{\tau}_{mn}(\tau))_q} \cdot \widehat{f}(m, n, \tau, \lambda(m, n))_q \right| d\tau \\
& \leq \int_0^{\bar{t}} \left| \frac{g_m^r(\widehat{\tau}_{mn}(\tau))}{\sum_{\{r' \in \mathcal{R} \mid \text{for some } i \pi_i(r')=m, \pi_{i+1}(r')=n\}} g_m^{r'}(\widehat{\tau}_{mn}(\tau))} \right. \\
& \quad \left. - \frac{g_m^r(\widehat{\tau}_{mn}(\tau)_q)_q}{\sum_{\{r' \in \mathcal{R} \mid \text{for some } i \pi_i(r')=m, \pi_{i+1}(r')=n\}} g_m^{r'}(\widehat{\tau}_{mn}(\tau)_q)_q} \right| \cdot \bar{f} \\
& \quad + \left| \widehat{f}(m, n, \tau, \lambda(m, n)) - \widehat{f}(m, n, \tau, \lambda(m, n))_q \right| d\tau \\
& = \bar{f} \cdot \int_0^{\bar{t}} \left| \frac{g_m^r(\widehat{\tau}_{mn}(\tau))}{\sum_{\{r' \in \mathcal{R} \mid \text{for some } i \pi_i(r')=m, \pi_{i+1}(r')=n\}} g_m^{r'}(\widehat{\tau}_{mn}(\tau))} \right. \\
& \quad \left. - \frac{g_m^r(\widehat{\tau}_{mn}(\tau)_q)_q}{\sum_{\{r' \in \mathcal{R} \mid \text{for some } i \pi_i(r')=m, \pi_{i+1}(r')=n\}} g_m^{r'}(\widehat{\tau}_{mn}(\tau)_q)_q} \right| d\tau \\
& \quad + \int_0^{\bar{t}} \left| \widehat{f}(m, n, \tau, \lambda(m, n)) - \widehat{f}(m, n, \tau, \lambda(m, n))_q \right| d\tau
\end{aligned}$$

We consider each of the two terms separately. For the first term, note that $\dot{\delta}$ is bounded above by $v(0, x(m, n)) < \infty$, so $d\widehat{\tau}_{mn}(\tau)_q/d\tau$ and $d\widehat{\tau}_{mn}(\tau)/d\tau$ are both bounded away from 0 by $1/v(0, x(m, n))$. Hence, sets of measure 0 in time τ are mapped to sets of measure 0 in the images of $\widehat{\tau}_{mn}(\cdot)_q$ and $\widehat{\tau}_{mn}$. Using Ascoli's theorem and passing to a subsequence if necessary, $\widehat{\tau}_{mn}(\cdot)_q \rightarrow \widehat{\tau}_{mn}(\cdot)$ uniformly. For if not, then $\lim_{q \rightarrow \infty} \widehat{\tau}_{mn}(\cdot)_q \neq \widehat{\tau}_{mn}(\cdot)$,

and there are two solutions to the differential equation (10), a contradiction. Since $\lim_{q \rightarrow \infty} g_m^r(\cdot)_q = g_m^r(\cdot)$ in L^1 norm, the convergence is a.s. Hence $g_m^r(\widehat{\tau}_{mn}(\cdot)_q) \rightarrow g_m^r(\widehat{\tau}_{mn}(\cdot))$ a.s. By Lebesgue's dominated convergence theorem, the first term converges to 0.

For the second term, recall that $\widehat{f}(m, n, \cdot, \lambda(m, n))$ is defined uniquely.

Now suppose that $\lim_{q \rightarrow \infty} \widehat{f}(m, n, \cdot, \lambda(m, n))_q \neq \widehat{f}(m, n, \cdot, \lambda(m, n))$. Then by Helly's theorem, we can find a subsequence of $\{f(m, n, \cdot, \lambda(m, n))_q\}_{q=1}^\infty$ converging to, say, $\widetilde{f}(m, n, \cdot, \lambda(m, n)) \neq \widehat{f}(m, n, \cdot, \lambda(m, n))$, where convergence is pointwise. By a uniqueness argument given above, it must be that $\widehat{f}(m, n, \cdot, \lambda(m, n))$ is not the exit density for a solution. From above,

$$\begin{aligned} & \int_0^{\lambda(m, n)} \int_0^{\bar{t}} \left| \widetilde{f}(m, n, t, \Delta) - \widehat{f}(m, n, t, \Delta) \right| \psi(\Delta) \varphi'(t) \\ & + \text{sign} \left(\widetilde{f}(m, n, t, \Delta) - \widehat{f}(m, n, t, \Delta) \right) \cdot \left(\Phi \left(\widetilde{f}(m, n, t, \Delta) \right) - \Phi \left(\widehat{f}(m, n, t, \Delta) \right) \right) \cdot \psi'(\Delta) \varphi(t) d\Delta dt \\ & \geq 0 \end{aligned}$$

For φ approximating the indicator function of $[0, \bar{t}]$, we have:

$$\begin{aligned} & \limsup_{t \rightarrow 0} \int_0^{\lambda(m, n)} \left| \widetilde{f}(m, n, t, \Delta) - \widehat{f}(m, n, t, \Delta) \right| \psi(\Delta) dt \\ & - \liminf_{t \rightarrow \bar{t}} \int_0^{\lambda(m, n)} \left| \widetilde{f}(m, n, t, \Delta) - \widehat{f}(m, n, t, \Delta) \right| \psi(\Delta) dt \\ & \geq - \int_0^{\lambda(m, n)} \int_0^{\bar{t}} \text{sign} \left(\widetilde{f}(m, n, t, \Delta) - \widehat{f}(m, n, t, \Delta) \right) \cdot \\ & \left(\Phi \left(\widetilde{f}(m, n, t, \Delta) \right) - \Phi \left(\widehat{f}(m, n, t, \Delta) \right) \right) \cdot \psi'(\Delta) d\Delta dt \end{aligned}$$

Taking ψ to approximate the indicator function of $[0, \lambda(m, n)]$,

$$\begin{aligned} & \geq \lim_{\Delta \rightarrow \lambda(m, n)} \sup_{\Delta \rightarrow \lambda(m, n)} \int_0^{\bar{t}} \text{sign} \left(\widetilde{f}(m, n, t, \Delta) - \widehat{f}(m, n, t, \Delta) \right) \cdot \\ & \left(\Phi \left(\widetilde{f}(m, n, t, \Delta) \right) - \Phi \left(\widehat{f}(m, n, t, \Delta) \right) \right) dt \\ & - \lim_{\Delta \rightarrow 0} \inf_{\Delta \rightarrow 0} \int_0^{\bar{t}} \text{sign} \left(\widetilde{f}(m, n, t, \Delta) - \widehat{f}(m, n, t, \Delta) \right) \cdot \\ & \left(\Phi \left(\widetilde{f}(m, n, t, \Delta) \right) - \Phi \left(\widehat{f}(m, n, t, \Delta) \right) \right) dt \end{aligned}$$

In sum, we have:

$$\begin{aligned}
& \limsup_{t \rightarrow 0} \int_0^{\lambda(m,n)} \left| \tilde{f}(m, n, t, \Delta) - \hat{f}(m, n, t, \Delta) \right| \psi(\Delta) dt - \\
& \liminf_{t \rightarrow \bar{t}} \int_0^{\lambda(m,n)} \left| \tilde{f}(m, n, t, \Delta) - \hat{f}(m, n, t, \Delta) \right| \psi(\Delta) dt \\
& \geq \lim_{\Delta \rightarrow \lambda(m,n)} \sup \int_0^{\bar{t}} \text{sign} \left(\tilde{f}(m, n, t, \Delta) - \hat{f}(m, n, t, \Delta) \right) \cdot \\
& \quad \left(\Phi \left(\tilde{f}(m, n, t, \Delta) \right) - \Phi \left(\hat{f}(m, n, t, \Delta) \right) \right) dt \\
& \quad - \lim_{\Delta \rightarrow 0} \inf \int_0^{\bar{t}} \text{sign} \left(\tilde{f}(m, n, t, \Delta) - \hat{f}(m, n, t, \Delta) \right) \cdot \\
& \quad \left(\Phi \left(\tilde{f}(m, n, t, \Delta) \right) - \Phi \left(\hat{f}(m, n, t, \Delta) \right) \right) dt
\end{aligned}$$

Since the left hand side (the first two terms) are zero, we obtain:

$$\begin{aligned}
& \lim_{\Delta \rightarrow 0} \inf \int_0^{\bar{t}} \text{sign} \left(\tilde{f}(m, n, t, \Delta) - \hat{f}(m, n, t, \Delta) \right) \cdot \\
& \quad \left(\Phi \left(\tilde{f}(m, n, t, \Delta) \right) - \Phi \left(\hat{f}(m, n, t, \Delta) \right) \right) dt \\
& \geq \lim_{\Delta \rightarrow \lambda(m,n)} \sup \int_0^{\bar{t}} \text{sign} \left(\tilde{f}(m, n, t, \Delta) - \hat{f}(m, n, t, \Delta) \right) \cdot \\
& \quad \left(\Phi \left(\tilde{f}(m, n, t, \Delta) \right) - \Phi \left(\hat{f}(m, n, t, \Delta) \right) \right) dt
\end{aligned}$$

As in Strub and Bayen (2006, p. 558), the left hand side is 0. The right hand side is non-negative (recall that flux Φ is increasing in density). Hence,

$$\begin{aligned}
& \lim_{\Delta \rightarrow \lambda(m,n)} \sup \int_0^{\bar{t}} \text{sign} \left(\tilde{f}(m, n, t, \Delta) - \hat{f}(m, n, t, \Delta) \right) \cdot \left(\Phi \left(\tilde{f}(m, n, t, \Delta) \right) - \Phi \left(\hat{f}(m, n, t, \Delta) \right) \right) dt \\
& \quad = 0
\end{aligned}$$

Now since Φ is strictly increasing in f , then we know that

$$\lim_{\Delta \rightarrow \lambda(m,n)} \sup \left\| \tilde{f}(m, n, \cdot, \Delta) - \hat{f}(m, n, \cdot, \Delta) \right\|_{L^1} = 0,$$

implying that $\tilde{f}(m, n, \cdot, \lambda(m, n)) = \hat{f}(m, n, \cdot, \lambda(m, n))$ a.s. (t) . Both $\tilde{f}(m, n, \cdot, \lambda(m, n))$ and $\hat{f}(m, n, \cdot, \lambda(m, n))$ are of bounded variation, so by Lemma 2.1 and Remark 2.1 of Bressan (2000), by taking right continuous versions, they are in fact equal, a contradiction.

Next, apply Schauder's theorem to the space \mathcal{G} with the L^1 norm and the mapping \mathcal{T} . This yields existence of at least one fixed point. To show that it is

unique, find the earliest time at which the two solutions diverge. Observe that for given initial conditions, behavior within a link is well-defined. So if two solutions exist and the earliest divergence between them occurs within a link, we have a contradiction. Thus, the divergence must occur at a node. Finding the earliest time at which such a divergence occurs, the initial conditions must be ill-defined, a contradiction.

■

5.2 Proof of Theorem 2

Proof: A mixed strategy is a measurable map $y : C \rightarrow [0, 1]^{|R| \times (T/\bar{\tau}-1)}$. We use the notation y^i to denote a vector component of y , so we impose the obvious condition $\sum_{i=1}^{|R| \times (T/\bar{\tau}-1)} y^i(c) = 1$ almost surely in c .

First, we can define a *strategy distribution* as $\int_C y \equiv \prod_{i=1}^{|R| \times (T/\bar{\tau}-1)} \int_C y^i(c) d\mu$. Second, we notice that the proof of Theorem 1 does not use the exact dynamic commuting route structure, but rather the strategy distribution induced by a dynamic route structure. In other words, the proof of Theorem 1 implies that for any given strategy distribution, there exists a unique pattern of traffic flows. Information about which commuter plays each strategy is irrelevant.

Third, we define the utility of a commuter for a mixed strategy and any strategy distribution. Fix $c \in C$. The utility function $u(c; \tau^d, l, R)$ was given in (9). For pure strategy i corresponding to $l(c), R(c), \tau^d(c)$, this is written as $\hat{u}^i(c, \int_C y) = u(c; \tau^d, l, R)$. We have argued that in the end the flows depend only on the strategy distribution. For technical reasons, it is useful here to define $u^i(c, \int_C y) \equiv -\infty$ if $\pi_1(R(c)) \neq O(c)$ or $\pi_{l(c)}(R(c)) \neq D(c)$; utility was undefined for this circumstance. Then for commuter $c \in C$, we can write the utility from the use of pure strategy i (a route and time of departure) given an aggregate strategy profile $\int_C y$, as $\hat{u}^i(c, \int_C y)$ and $\hat{u}(c, y) = \prod_{i=1}^{|R| \times (T/\bar{\tau}-1)} \hat{u}^i(c, \int_C y)$, where the dynamic route structure (τ^d, l, R) generates the strategy distribution $\int_C y$. For this to be well-defined, we are using the fact that the utility will depend only on the strategy distribution generated by the dynamic commuting route structure, and the fact that this can now be chosen arbitrarily since we no longer stick to the requirement that the origin and destination nodes are pre-specified. Finally, we can define the utility of commuter c from using mixed strategy $y(c)$ by $y(c) \cdot \hat{u}(c, y)$.

It is clear from this set of definitions that our model satisfies two of the assumptions of Schmeidler (1973), namely the measurability assumption (b)

and the fact that utility depends only on the strategy distribution, not on individual strategies. Assumption (a), regarding the continuity of v in its second argument, remains to be verified.

We take a sequence of mixed strategies $\{y_q\}_{q=1}^\infty$ such that $\lim_{q \rightarrow \infty} y_q = y$ in the L^1 weak topology, and prove that for each $c \in C$, $\lim_{q \rightarrow \infty} \hat{u}(c, y_q) = \hat{u}(c, y)$. Our hypothesis implies $\lim_{q \rightarrow \infty} \int_C y_q = \int_C y$. Let $g \in \mathcal{G}$ be the fixed point associated with the initial conditions y ,³⁶ and let $g_q \in \mathcal{G}$ be the fixed point associated with the initial conditions y_q . Thus, we have an associated sequence $\{g(\cdot)_q\}_{q=1}^\infty \subseteq \mathcal{G}$ where for each q , $g(\cdot)_q = \mathcal{T}(g(\cdot)_q)$. Since \mathcal{G} is compact, there is a converging subsequence. Now pass to *any* converging subsequence, call it $\{g(\cdot)_{q_p}\}_{p=1}^\infty \subseteq \mathcal{G}$, where $\lim_{p \rightarrow \infty} g(\cdot)_{q_p} = \hat{g}(\cdot)$. By continuity of \mathcal{T} , $\hat{g}(\cdot) = \mathcal{T}(\hat{g}(\cdot))$. Hence, $\hat{g} = g$, and $\lim_{q \rightarrow \infty} g_q = g$. We use an analogous argument below for both flows and progress along a link.

Define

$$\begin{aligned} \mathcal{F}_{mn} \equiv & \left\{ \hat{f}(m, n, \cdot, \cdot) \text{ measurable on } [0, \bar{t}] \times [0, \lambda(m, n)] \right. \\ & \left. | 0 \leq \hat{f}(m, n, \cdot, \cdot) \leq \bar{f} \text{ a.s., } \hat{f}(m, n, 0, \Delta) = 0 \ \forall \ \Delta \geq 0 \right\} \end{aligned}$$

Then we can define:

$$\mathcal{F} \equiv \prod_{m, n=1, m \neq n}^N \mathcal{F}_{mn}$$

We denote a typical element of \mathcal{F} by $\hat{f} = \left(\hat{f}(m, n, \cdot, \cdot) \right)_{m, n=1, m \neq n}^N$.

Now for each q there exists a unique solution $\hat{f}_q \in \mathcal{F}$ associated with initial conditions g_q . There is also a unique flow $\hat{f} \in \mathcal{F}$ associated with g . Impose the weak* topology on the flows as a subset of L^∞ . Applying the Banach-Alaoglu theorem, there is a converging subsequence. Now pass to any converging subsequence, call it $\left\{ \hat{f}(\cdot)_{q_p} \right\}_{p=1}^\infty \subseteq \mathcal{F}$, where $\lim_{p \rightarrow \infty} \hat{f}(\cdot)_{q_p} = \tilde{f}(\cdot)$, where convergence is pointwise a.s. in (τ, Δ) . As in the proof of Theorem 1, it must be that $\hat{f} = \tilde{f}$.

Now for each q there exists a unique solution $\delta_q \in \mathcal{D}$ associated with flow \hat{f}_q . There is also a unique solution $\delta \in \mathcal{D}$ associated with \hat{f} . Impose the uniform topology on the solutions as a subset of C_0 . Applying Ascoli's theorem, there is a converging subsequence. Now pass to any converging subsequence, call it $\left\{ \delta(\cdot)_{q_p} \right\}_{p=1}^\infty \subseteq \mathcal{D}$, where $\lim_{p \rightarrow \infty} \delta(\cdot)_{q_p} = \hat{\delta}(\cdot)$.

³⁶ Although y represents a mixed strategy, as we have noted, all that matters is the distribution of initial conditions (namely the distribution over routes and departure times), so flows and arrival times can be found uniquely for each mixed strategy profile using Theorem 1.

Next define $\widehat{e}(m, n, \tau, \widehat{\tau})_{q_p} \equiv \widehat{f}(m, n, \tau, \delta_{mn}(\widehat{\tau}, \tau)_{q_p})_{q_p}$. Now

$$TV \left(\widehat{f}(m, n, \tau, \delta_{mn}(\widehat{\tau}, \tau)_{q_p})_{q_p} \right) \leq TV \left(\widehat{f}(m, n, \tau, 0)_{q_p} \right) \leq |\mathcal{R}| \cdot (T/\bar{\tau} + 1) \cdot \bar{f}$$

So applying Helly's theorem and passing to a further subsequence if necessary,

$$\lim_{q \rightarrow \infty} \widehat{e}(m, n, \tau, \widehat{\tau})_{q_p} = \widehat{e}(m, n, \tau, \widehat{\tau})$$

where convergence is pointwise in τ and $TV(\widehat{e}(m, n, \tau, \widehat{\tau})) \leq |\mathcal{R}| \cdot (T/\bar{\tau} + 1) \cdot \bar{f}$.

So for each p ,

$$\frac{\partial \delta_{mn}(\widehat{\tau}, \tau)_{q_p}}{\partial \tau} = v \left(f(m, n, \tau, \delta(\widehat{\tau}, \tau)_{q_p})_{q_p}, x_{mn} \right) = v \left(\widehat{e}(m, n, \tau, \widehat{\tau})_{q_p}, x_{mn} \right)$$

so

$$\lim_{p \rightarrow \infty} \frac{\partial \delta_{mn}(\widehat{\tau}, \tau)_{q_p}}{\partial \tau} = v \left(\widehat{e}(m, n, \tau, \widehat{\tau}), x_{mn} \right)$$

Next suppose that $\lim_{p \rightarrow \infty} \frac{\partial \delta(\widehat{\tau}, \tau)_{q_p}}{\partial \tau} \neq \frac{\partial \widehat{\delta}(\widehat{\tau}, \tau)}{\partial \tau}$ on a set of positive measure in τ . Hence, by the fundamental theorem of calculus and Lebesgue's dominated convergence theorem, there exists τ' such that

$$\begin{aligned} \widehat{\delta}_{mn}(\widehat{\tau}, \tau') &= \int_0^{\tau'} \frac{\partial \widehat{\delta}_{mn}(\widehat{\tau}, \tau)}{\partial \tau} d\tau = \int_0^{\tau'} \lim_{p \rightarrow \infty} \frac{\partial \delta_{mn}(\widehat{\tau}, \tau)_{q_p}}{\partial \tau} d\tau \\ &= \lim_{p \rightarrow \infty} \int_0^{\tau'} \frac{\partial \delta_{mn}(\widehat{\tau}, \tau)_{q_p}}{\partial \tau} d\tau \neq \int_0^{\tau'} \frac{\partial \widehat{\delta}_{mn}(\widehat{\tau}, \tau)}{\partial \tau} d\tau = \widehat{\delta}_{mn}(\widehat{\tau}, \tau') \end{aligned}$$

This is obviously a contradiction. So $\lim_{p \rightarrow \infty} \frac{\partial \delta_{mn}(\widehat{\tau}, \tau)_{q_p}}{\partial \tau} = \frac{\partial \widehat{\delta}_{mn}(\widehat{\tau}, \tau)}{\partial \tau}$ a.s. From (6) we know that $\delta_{mn}(\widehat{\tau}, \widehat{\tau}) = \widehat{\delta}_{mn}(\widehat{\tau}, \widehat{\tau}) = 0$, so by integrating, $\widehat{\delta}(\cdot) = \delta(\cdot)$.

Fix a route r and a departure time $\widehat{\tau}$. Define

$$\begin{aligned} \tau_{mn}^*(\widehat{\tau}) &\equiv \min \{0 \leq \tau \leq \bar{t} \mid \delta_{mn}(\widehat{\tau}, \tau) = \lambda(m, n)\} \\ &= \delta_{mn}^{-1}(\lambda(m, n))(\widehat{\tau}) \end{aligned}$$

Now let τ^d and $\tau^{d'}$ be origin departure time choices for route r , and let τ and τ' be associated perturbations, where $\widehat{\tau} = \tau^d + \tau$ and $\widehat{\tau}' = \tau^{d'} + \tau'$. Then

$$\left| \tau_{mn}^*(\widehat{\tau}) - \tau_{mn}^*(\widehat{\tau}') \right| \leq \frac{\lambda(m, n)}{v(\bar{f}, x_{mn})} \cdot |\widehat{\tau} - \widehat{\tau}'|$$

Thus, arrival time at the final destination can be written as: $\widehat{\tau}_l(\tau^d + \tau') = \tau_{\pi(l-1), \pi(l)}^* \left(\tau_{\pi(l-2), \pi(l-1)}^* \left(\cdots \tau_{\pi(1), \pi(2)}^* (\tau^d + \tau') \cdots \right) \right)$. Hence,

$$\left| \widehat{\tau}_l(\tau^d + \tau') - \widehat{\tau}_l(\tau^{d'} + \tau') \right| \leq \prod_{i=1}^{l-1} \frac{\lambda(\pi(i), \pi(i+1))}{v(\bar{f}, x_{\pi(i)\pi(i+1)})} \cdot |\tau^d + \tau - \tau^{d'} - \tau'|$$

Define

$$\begin{aligned} \Upsilon_r \equiv & \left\{ \hat{\tau}_l : [0, T] \rightarrow [0, \bar{t}] \text{ measurable} \mid |\hat{\tau}_l(\tau^d + \tau) - \hat{\tau}_l(\tau^{d'} + \tau')| \right. \\ & \left. \leq \prod_{i=1}^{l-1} \frac{\lambda(\pi(i), \pi(i+1))}{v(\bar{f}, x_{\pi(i)\pi(i+1)})} \cdot |\tau^d + \tau - \tau^{d'} - \tau'| \right\} \end{aligned}$$

By Ascoli's theorem, Υ_r is a compact subset of C_0 .

For each q there is a unique $\delta(\cdot)_q$ and thus a unique $\hat{\tau}_l(\cdot)_q \in \Upsilon_r$. There is also a unique $\hat{\tau}_l(\cdot) \in \Upsilon_r$ associated with $\delta(\cdot)$. So there is a converging subsequence associated with $\{\delta(\cdot)_q\}_{q=1}^\infty$. Now take any converging subsequence of $\{\hat{\tau}_l(\cdot)_q\}_{q=1}^\infty$, call it $\{\hat{\tau}_l(\cdot)_{q_p}\}_{p=1}^\infty$. It has a limit: $\hat{\tau}'_l(\cdot) \in \Upsilon_r$. Suppose that $\hat{\tau}'_l(\cdot) \neq \hat{\tau}_l(\cdot)$. Now since $\{\delta(\cdot)_{q_p}\}_{p=1}^\infty$ converges uniformly to $\delta(\cdot)$, for each $\tau^d + \tau' \in [0, T]$, $\lim_{p \rightarrow \infty} \hat{\tau}_l(\tau^d + \tau')_{q_p} = \hat{\tau}_l(\tau^d + \tau')$, so in fact $\hat{\tau}'_l(\cdot) = \hat{\tau}_l(\cdot)$, a contradiction.

Apply Schmeidler (1973), theorems 1 and 2, there exists a Nash equilibrium in pure strategies.

Finally, consider the case where Φ is non-decreasing (instead of strictly increasing) in f , and as always $\Phi(f) \equiv v(f) \cdot f$. Let $\hat{v}(f) = v(f) + \epsilon$, where $\epsilon > 0$ is small. Then since $v(f)$ is non-increasing in f , so is $\hat{v}(f)$. Moreover, $\hat{\Phi}(f) \equiv \hat{v}(f) \cdot f = (v(f) + \epsilon) \cdot f = \Phi(f) + \epsilon \cdot f$, so $\hat{\Phi}(f)$ is strictly increasing in f . Apply our results to the modified game using $\hat{v}(f)$ and $\hat{\Phi}(f)$ to obtain an equilibrium in pure strategies for each ϵ . As the number of strategies is actually finite, we can find an accumulation point of the strategy profile as $\epsilon \rightarrow 0$. Using continuity of the payoffs (as demonstrated above), by a standard argument the accumulation point is an equilibrium profile for $\epsilon = 0$.

■

5.3 Proof of Theorem 3

Proof: Given an outward bound tree network, there is no route choice. The strategy profile we propose as a Pareto efficient Nash equilibrium is to distribute each type of commuter, where type is defined as an origin-destination pair, uniformly across all departure times. Clearly this is a Nash equilibrium, as all commuters of a given type have the same travel time and thus receive the same utility. Now suppose that there is a strategy profile that Pareto dominates the Nash equilibrium profile. Thus, it must be that there is some departure time that has a higher than average density. For this departure time, there is some type that has a higher than average density (where the

average is over departure times for this type). The commuters of this type with this departure time will have a longer commute than at Nash equilibrium, contradicting that the alternative strategy profile Pareto dominates the Nash equilibrium profile.

■